

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

QM protein - protein search, using sw model

Run on: June 29, 2004, 16:54:09 ; Search time 49.3653 Seconds
(without alignment)
1562.545 Million cell updates/sec

Title: US-09-894-912A-13

Perfect score: 1516

Sequence: 1 MGLRLISWLFILNFMEYI.....QQKKRKVDKQKSVSVSTVH 273

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq_29Jan04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1516	100.0	273	4	Aae13151 Human ste
2	1516	100.0	273	6	Abo44414 Human ste
3	1505	99.3	272	4	Aab99220 Human thr
4	1505	99.3	272	4	Aam78328 Human pro
5	1505	99.3	272	4	Aae13168 Human ste
6	1505	99.3	272	4	Aae13150 Human ste
7	1505	99.3	272	6	Abr62112 Human clo
8	1505	99.3	272	6	Abo44432 Human ste
9	1505	99.3	272	6	Abo44413 Human ste
10	1472	97.1	265	4	Aae13163 Human sec
11	1472	97.1	265	6	Abo44427 Human sec
12	1472	97.1	292	2	Aaw85607 Secreted
13	1472	97.1	292	4	Aae13170 Human SCR
14	1472	97.1	292	5	Abp61846 Human pol
15	1472	97.1	292	6	Abr62114 Human sec
16	1472	97.1	292	6	Abo44434 Human SCR
17	1452	95.8	292	4	Aab93875 Human pro
18	1400	92.3	251	4	Aae13153 Human mat
19	1400	92.3	251	6	Abo44417 Human ste
20	1310.5	86.4	279	4	Aae13167 Mouse ste
21	1310.5	86.4	279	6	Abo44431 Mouse ste
22	1178	77.7	239	6	Abu52396 Human GPC
23	1111	73.3	195	6	Abu52398 Human GPC
24	1107	73.0	195	6	Abu52397 Human GPC
25	903	59.6	160	4	Abb11374 Human sec

26	903	59.6	160	4	Aam79312	Human pro
27	903	59.6	160	4	Aae13149	Human ste
28	903	59.6	160	6	Abo44415	Human ste
29	656	43.3	263	6	Abr62108	Secreted
30	656	43.3	263	6	ABR62115	Secreted
31	656	43.3	263	6	ABR58489	Human sec
32	656	43.3	263	7	ADe07919	Novel pro
33	648.5	42.8	243	6	ABR62110	Secreted
34	644	42.5	265	6	ABR62113	Mouse thr
35	638	42.1	229	4	Aae13162	Mouse thr
36	638	42.1	229	6	Abo44426	Human sec
37	584.5	38.6	243	6	Aae37115	Human sec
38	577.5	38.1	243	5	ABG76508	DNA encod
39	577.5	38.1	243	6	ABR62106	Secreted
40	576.5	38.0	243	6	Aae36166	Mouse SCR
41	573	37.8	250	6	ABR62101	Secreted
42	545	35.9	222	6	ABR62107	Secreted
43	540.5	35.7	229	6	ABR62102	Secreted
44	472	31.1	234	7	ADB76146	Novel hum
45	276	18.2	46	4	Aae13155	Human SCR

ALIGNMENTS

RESULT 1
Aae13151
ID AAE13151 standard; protein; 273 AA.
XX
AC AAE13151;
XX
XX
DT 28-JAN-2002 (first entry)
XX
DE Human stem cell growth factor-like protein #3.
XX
KW Human; stem cell growth factor-like protein; antiinflammatory; nootropic;
KW neuroprotective; vulnery; cytostatic; anticonvulsant; immunostimulant;
KW vasotropic; virucide; dermatological; tranquiliser; cerebroprotective;
KW osteopathic; immunodeficiency syndrome; chronic granulomatous disease;
KW duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;
KW acquired immune deficiency syndrome; agammaglobulinaemia; thalassaemia;
KW Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;
KW adrenal white matter degeneration; anaemia; neurodegenerative disease;
KW Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;
KW severe combined immunodeficiency; immune disorder; autoimmune disorder;
KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
KW autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;
KW supporting factor for the proliferation of stem cell.

XX Homo sapiens.

XX WO200177169-A2.

XX 18-OCT-2001.

XX 05-APR-2001; 2001WO-US011208.

XX 05-APR-2000; 2000US-00543774.

XX 28-JUN-2000; 2000US-0215733P.

XX 09-JAN-2001; 2001US-00757562.

XX 05-FEB-2001; 2001US-0266614P.

XX (HYSE-) HYSEQ INC.

XX (KIRI) KIRIN BEER KK.

XX Tang TV, Labat I, Tillinghaast JS, Sinku A, Liu C, Drmanac RT;

XX Stache-Crain B, Dickson M, Mize NK, Nishikawa M;

XX WPI; 2001-657166/75.

XX N-PSDB; AAD21725.

XX Novel stem cell growth factor like polypeptides and polynucleotides for
PT identifying modulators useful for treating diseases such as Alzheimer's

PT disease, cancer, rheumatoid arthritis, osteoporosis.
 PS Claim 28; Page 214-215; 232pp; English.
 XX
 CC The patent discloses novel stem cell growth factor-like proteins and
 CC polynucleotides encoding them. Proteins of the invention are also known
 CC as supporting factor for the proliferation of stem cells (SCR-1). Stem
 CC cell growth factor-like proteins are useful for supporting proliferation
 CC or survival of a stem cell or germ cell which is preferably primordial
 CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem
 CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent
 CC cell. The haematopoietic progenitor cell cultured using stem cell growth
 CC factor-like proteins can replace as a graft for the bone marrow
 CC transplantation or cord blood transplantation for treating a variety of
 CC diseases such as immunodeficiency syndrome, chronic granulomatous
 CC disease, duplicated immunodeficiency syndrome, agammaglobulinaemia,
 CC Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),
 CC thalassaemia, haemolytic anaemia due to enzyme defect, congenital anaemia
 CC such as sickle cell anaemia, Gaucher's disease, lysosomal storage
 CC diseases such as mucopolysaccharidosis, adrenal white matter
 CC degeneration, a variety of cancer and tumours. Proteins of the invention
 CC are useful for treating diseases such as Parkinson's disease, Alzheimer's
 CC disease and other neurodegenerative diseases, thrombocytopaenia, immune
 CC deficiencies and disorders such as severe combined immunodeficiency
 CC (SCID) and autoimmune disorders such as multiple sclerosis, systemic
 CC lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary
 CC inflammation. Sequences of the invention are also useful in gene therapy.
 XX The present sequence is stem cell growth factor-like protein from human
 XX Sequence 273 AA;
 Query Match 100.0%; Score 1516; DB 4; Length 273;
 Best Local Similarity 100.0%; Pred. No. 3.8e-111; Indels 0; Gaps 0;
 Matches 273; Conservative 0; Mismatches 0;
 QY 1 MGHRLISWLFILFMFMEYIGSQNASRGRQRMRHNPVSGQGCGCATCDYNGCLSCRP 60
 DB 1 MGHRLISWLFILFMFMEYIGSQNASRGRQRMRHNPVSGQGCGCATCDYNGCLSCRP 60
 QY 61 RLFFALERIGMQIGVCLSSCSGYTYGTRYPDINKCTCKACDCTCFNKNFTCKSGFY 120
 DB 61 RLFFALERIGMQIGVCLSSCSGYTYGTRYPDINKCTCKACDCTCFNKNFTCKSGFY 120
 QY 121 LHLGKCLDNCPEGLANNHMCVSIHVCEVSEWNPSPCTKKGKTCGFRGTETRVREI 180
 DB 121 LHLGKCLDNCPEGLANNHMCVSIHVCEVSEWNPSPCTKKGKTCGFRGTETRVREI 180
 QY 181 IQHPSAKGNLCPTTNETRCKTVQRKKCKGKGRKRRKPKKPNKGBSKBAIPDSKSLE 240
 DB 181 IQHPSAKGNLCPTTNETRCKTVQRKKCKGKGRKRRKPKKPNKGBSKBAIPDSKSLE 240
 QY 241 SSK3IPEQRNKKQKQKRVQDKQKSVSVSTVH 273
 DB 241 SSK3IPEQRNKKQKQKRVQDKQKSVSVSTVH 273
 RESULT 2
 ABO44414
 ID ABO44414 standard; protein; 273 AA.
 XX ABO44414;
 AC ABO44414;
 XX
 DT 30-SEP-2003 (first entry)
 XX
 DE Human stem cell growth factor-like protein, SCR 1 #2.
 KW Stem cell growth factor-like protein; antianemic; anti-HIV; SCR-1;
 KW immunostimulant; vulnery; haematopoietic stem cell; gene therapy;
 KW supporting factor for proliferation of stem cells; wound healing;
 KW haematopoietic progenitor cell; stromal cell; AIDS; thalassaemia;
 KW bone marrow transplantation; cord blood transplantation;
 KW chronic granulomatous disease; duplicated immunodeficiency syndrome;
 KW agammaglobulinaemia; Wiskott-Aldrich syndrome; haemolytic anaemia;
 KW congenital anaemia; sicklaemia; Gaucher's disease; morphogenesis;
 KW epithelial cell growth; ovarian follicle development; nerve cell growth;
 KW cartilage remodeling; bone growth; immunosuppression; human.
 XX Homo sapiens.
 OS US2003044792-A1.
 XX 06-MAR-2003.
 PD 28-JUN-2001; 2001US-00894912.
 PF 28-JUN-2000; 2000US-0215733P.
 PR 05-FEB-2001; 2001US-0266614P.
 PR 05-APR-2001; 2001US-0282397P.
 XX (TANG/) TANG Y T.
 PA (LABA/) LABAT I.
 PA (DRMA/) DRMANAC R T.
 PA (MIZE/) MIZE N.
 PA (NISH/) NISHIKAWA M.
 PA (CHAO/) CHAO C.
 XX Tang YT, Labat I, Drmanac RT, Mize N, Nishikawa M, Chao C;
 WPI: 2003-625403/59.
 DR N-PSDB; ACH04325.
 XX Novel isolated polypeptide having stem cell growth factor activity,
 PT useful for promoting wound healing, and as a medicine to proliferate or
 PT support human hematopoietic stem cells or human hematopoietic progenitor
 PT cells.
 PT Claim 23; Page 72-73; 96pp; English.
 XX The invention relates to an isolated stem cell growth factor-like
 CC polypeptide (referred as supporting factor for proliferation of stem
 CC cells (SCR-1)) from mouse or human, or its mature protein portion, or
 CC fragment, analogue, variant or derivative, that retains stem cell growth
 CC factor activity. Also included are an isolated polynucleotide encoding
 CC SCR-1 (or its mature protein portion, or fragment, analogue, variant or
 CC derivative, that retains stem cell growth factor activity, or the
 CC complement of the polynucleotide), an (expression) vector comprising the
 CC SCR-1 polynucleotide, a host cell genetically engineered to contain the
 CC SCR-1 polynucleotide in operative association with a regulatory sequence
 CC that controls expression of the polynucleotide in the host cell,
 CC preparation of the SCR-1 polypeptide, a polypeptide which is an
 CC expression product of the SCR-1 polynucleotide (the polypeptide having an
 CC activity to support proliferation or survival of haematopoietic stem cell
 CC or haematopoietic progenitor cell, with a proviso that C-terminal aa
 CC sequence does not comprise the aa sequence appearing as ABO44433), an
 CC isolated SCR-1 polypeptide with stem cell growth factor activity and
 CC lacking any 10 consecutive aas from ABO44430, an isolated polypeptide
 CC with stem cell growth factor activity having at least an aa sequence
 CC appearing as ABO44428 and ABO44429, a culture medium comprising the SCR-1
 CC polypeptide to maintain survival of or promote proliferation of a stem
 CC cell or germ cell, an anti-SCR-1 antibody, a nucleic acid array
 CC comprising the SCR-1 polynucleotide or a unique segment of the SCR-1
 CC polynucleotide attached to a surface, a stromal cell genetically
 CC engineered to express the SCR-1 polypeptide to support proliferation or
 CC survival of a stem cell or germ cell and an implant comprising a cell
 CC genetically engineered to express the SCR-1 polypeptide to support
 CC proliferation or survival of a stem cell or germ cell. The SCR-1
 CC polypeptide is useful for identifying a compound that binds to the SCR-1
 CC polypeptide and for maintaining survival of or promoting proliferation of
 CC a stem cell, a germ cell, a haematopoietic stem cell or a haematopoietic
 CC progenitor cell. The SCR-1 polypeptide is useful for promoting wound
 CC healing. The human haematopoietic stem cell or human haematopoietic
 CC progenitor cell culture using the SCR-1 polypeptide can replace as a
 CC graft for the conventional bone marrow transplantation or cord blood
 CC transplantation. The transplantation of haematopoietic stem cells can be
 CC employed as a therapy for treating diseases such as chronic granulomatous
 CC diseases, duplicated immunodeficiency syndrome, agammaglobulinaemia,

CC Wiskott-Aldrich syndrome, AIDS, etc., thalassemia, haemolytic anaemia
 CC due to enzyme defect, congenital anaemia such as sicklaemia, Gaucher's
 CC disease etc. the SCR-1 polypeptide is useful for cell growth and
 CC morphogenesis, including tissue specific stem cell growth, epithelial
 CC cell growth and regulation, ovarian follicle development, promoting nerve
 CC cell growth, sustaining neuronal populations, cartilage remodeling, bone
 CC growth and immunosuppression. The present sequence is a Human SCR-1
 CC protein
 XX
 SQ

Sequence 273 AA;

Query Match 100.0%; Score 1516; DB 6; Length 273;
 Best Local Similarity 100.0%; Pred. No. 3.8e-111;
 Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLRLISLWLFILNFMFMEYIGSNASGRQRMRHNPVSGCGGCATCSDYNGCLSCKP 60
 DB 1 MGLRLISLWLFILNFMFMEYIGSNASGRQRMRHNPVSGCGGCATCSDYNGCLSCKP 60

QY 61 RLFFALERIGMKQIGVCLSSCPGYYGTRYPDINKTKCKADCDTCFNKNFCTCKSGFY 120
 DB 61 RLFFALERIGMKQIGVCLSSCPGYYGTRYPDINKTKCKADCDTCFNKNFCTCKSGFY 120

QY 121 LHGLKLDNCPGLEANNHTMECVSIHVCEVSWNPSPCTKKGKTCGPKRGTTETRVREI 180
 DB 121 LHGLKLDNCPGLEANNHTMECVSIHVCEVSWNPSPCTKKGKTCGPKRGTTETRVREI 180

QY 181 IQHPSAGNLCPPTNETRKTCTVQRKKCKGKGRKRKRKPKNGESKEAIPDSKSL 240
 DB 181 IQHPSAGNLCPPTNETRKTCTVQRKKCKGKGRKRKRKPKNGESKEAIPDSKSL 240

QY 241 SSKEIPEQRNKKQKKRKVDKQKSVSVTVH 273
 DB 241 SSKEIPEQRNKKQKKRKVDKQKSVSVTVH 273

RESULT 3

AAB99220
 ID AAB99220 standard; protein; 272 AA.

AC AAB99220;

DT 07-SEP-2001 (first entry)

XX Human thrombospondin-30.

XX Human; thrombospondin-30; cytostatic; anti-HIV; antiinflammatory;
 KW malignant tumour; haemopathy; HIV infection; immunological disease;
 KW inflammation disease.

XX Homo sapiens.

XX WO200140294-A1.

XX 07-JUN-2001.

XX 27-NOV-2000; 2000WO-CN000494.

XX 29-NOV-1999; 99CN-00124148.

XX (BIOR-) BIORAD GENE DEV LTD SHANGHAI.

XX Mao Y, Xie Y;

XX WPI; 2001-397948/42.

XX N-PSDB; AA045131.

XX Human thrombospondin-30 and polynucleotide is useful in diagnosis and
 PT treatment of, e.g., malignant tumor, hemopathy, HIV infection,
 PT immunological diseases and various inflammatory diseases.

XX Claim 1; Page 27-28; 33pp; Chinese.

CC The present sequence is the protein sequence for human thrombospondin-30.
 CC Thrombospondin-30 protein and coding sequence are useful in the diagnosis
 CC and treatment of malignant tumour, haemopathy, HIV infection,
 CC immunological diseases and various inflammation diseases. In addition
 CC thrombospondin-30 protein may be used for screening mimics, agonists,
 CC antagonists or inhibitors, or for use in peptide fingerprinting
 CC identification. The thrombospondin-30 coding sequence may be used as
 CC primers for nucleic acid amplification reaction or as probes for
 CC hybridisation reaction, or in producing gene chips or microarrays
 XX

Sequence 272 AA;

Query Match 99.3%; Score 1505; DB 4; Length 272;
 Best Local Similarity 100.0%; Pred. No. 2.8e-110;
 Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HLRLISLWLFILNFMFMEYIGSNASGRQRMRHNPVSGCGGCATCSDYNGCLSCKPRL 62
 DB 2 HLRLISLWLFILNFMFMEYIGSNASGRQRMRHNPVSGCGGCATCSDYNGCLSCKPRL 61

QY 63 PFALERIGMKQIGVCLSSCPGYYGTRYPDINKTKCKADCDTCFNKNFCTCKSGFY 122
 DB 62 PFALERIGMKQIGVCLSSCPGYYGTRYPDINKTKCKADCDTCFNKNFCTCKSGFY 121

QY 123 LGKLDNCPGLEANNHTMECVSIHVCEVSWNPSPCTKKGKTCGPKRGTTETRVREI 182
 DB 122 LGKLDNCPGLEANNHTMECVSIHVCEVSWNPSPCTKKGKTCGPKRGTTETRVREI 181

QY 183 HPSAGNLCPPTNETRKTCTVQRKKCKGKGRKRKRKPKNGESKEAIPDSKSL 242
 DB 182 HPSAGNLCPPTNETRKTCTVQRKKCKGKGRKRKRKPKNGESKEAIPDSKSL 241

QY 243 KEIPEQRNKKQKKRKVDKQKSVSVTVH 273
 DB 242 KEIPEQRNKKQKKRKVDKQKSVSVTVH 272

RESULT 4

AAM78328
 ID AAM78328 standard; protein; 272 AA.

AC AAM78328;

DT 06-NOV-2001 (first entry)

XX Human protein SEQ ID NO 990.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.

XX Homo sapiens.

XX WO200157190-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US004098.

XX 03-FEB-2000; 2000US-00496914.

XX 27-APR-2000; 2000US-00568875.

XX 19-JUL-2000; 2000US-00620325.

XX 01-SEP-2000; 2000US-00654936.

XX 15-SEP-2000; 2000US-00663561.

XX 20-OCT-2000; 2000US-00693325.

XX 30-NOV-2000; 2000US-00728422.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;

PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 DR WPI; 2001-476283/51.
 DR N-PSDB; AAK51461.
 XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.
 XX Claim 20; Page 3214-3215; 6221pp; English.
 PS The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoietic regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication
 XX Sequence 272 AA;
 SQ
 Query Match 99.3%; Score 1505; DB 4; Length 272;
 Best Local Similarity 100.0%; Pred. No. 2.8e-110;
 Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 HURLISWLFILNFMEYIGSQNASRRORRRHHPNVSQCGGCATCSDYNGCLSKPRL 62
 DB 2 HURLISWLFILNFMEYIGSQNASRRORRRHHPNVSQCGGCATCSDYNGCLSKPRL 61
 QY 63 PFALERIGMKQIGVCLSSPCSGYGYTRYPDINKTKCRADCDTCFNKNFCTKCKSGFYHL 122
 DB 62 PFALERIGMKQIGVCLSSPCSGYGYTRYPDINKTKCRADCDTCFNKNFCTKCKSGFYHL 121
 QY 123 LQKCLDNCPEGLAANNHTECVSIHVCEVSEVNPSPCTKCKTGFGRGTFRVRELIQ 182
 DB 122 LQKCLDNCPEGLAANNHTECVSIHVCEVSEVNPSPCTKCKTGFGRGTFRVRELIQ 181
 QY 183 HPSAKGNLCPTNTRKCTVQRKKQKGRGKGRKRRKKNKESKEAI PDSKLSRSS 242
 DB 182 HPSAKGNLCPTNTRKCTVQRKKQKGRGKGRKRRKKNKESKEAI PDSKLSRSS 241
 QY 243 KEIPQRENKQKQKRRKQVQDKQKSVSVTVH 273
 DB 242 KEIPQRENKQKQKRRKQVQDKQKSVSVTVH 272
 RESULT 5
 AAE13168
 ID AAE13168 standard; protein; 272 AA.
 AC AAE13168;
 XX 28-JAN-2002 (first entry)
 XX Human stem cell growth factor-like protein #4.
 DE Human; stem cell growth factor-like protein; antiinflammatory; neurotropic;
 KW neuroprotective; vulnary; cyrostatic; anticonvulsant; immunostimulant;
 KW vasotrophic; virucide; dermatological; tranquilliser; cerebroprotective;
 KW osteopathic; immunodeficiency syndrome; chronic granulomatous disease;
 KW acquired immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;
 KW Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;
 KW adrenal white matter degeneration; anaemia; neurodegenerative disease;
 KW Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;
 KW severe combined immunodeficiency; immune disorder; autoimmune disorder;
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
 KW autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;

KW supporting factor for the proliferation of stem cell.
 XX Homo sapiens.
 OS Key Location/Qualifiers
 FH Peptide 1..21
 FT Protein 22..272
 FT /note= "Human mature stem cell growth factor-like
 FT protein"
 XX WO200177169-A2.
 XX 18-OCT-2001.
 PD 05-APR-2001; 2001WO-US011208.
 XX 05-APR-2000; 2000US-00543774.
 XX 28-JUN-2000; 2000US-0215733P.
 PR 09-JAN-2001; 2001US-00757562.
 PR 05-FEB-2001; 2001US-0266614P.
 XX (HYSE-) HYSEQ INC.
 PA (KIRI) KIRIN BEER KK.
 XX Tang TY, Labat I, Tillinghaast JS, Sinku A, Liu C, Drmanac RT;
 PI Stache-Crain B, Dickson M, Mize NK, Nishikawa M;
 XX WPI; 2001-657166/75.
 DR N-PSDB; AAD21728.
 XX Novel stem cell growth factor like polypeptides and polynucleotides for
 PT identifying modulators useful for treating diseases such as Alzheimer's
 FT disease, cancer, rheumatoid arthritis, osteoporosis.
 XX Claim 28; Page 226-227; 232pp; English.
 XX The patent discloses novel stem cell growth factor-like proteins and
 CC polynucleotides encoding them. Proteins of the invention are also known
 CC as supporting factor for the proliferation of stem cells (SCR-1). Stem
 CC cell growth factor-like proteins are useful for supporting proliferation
 CC or survival of a stem cell or germ cell which is preferably primordial
 CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem
 CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent
 CC cell. The haematopoietic progenitor cell cultured using stem cell growth
 CC factor-like proteins can replace as a graft for the bone marrow
 CC transplantation or cord blood transplantation for treating a variety of
 CC diseases such as immunodeficiency syndrome, chronic granulomatous
 CC disease, duplicated immunodeficiency syndrome, agammaglobulinaemia,
 CC Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),
 CC thalassaemia, haemolytic anaemia due to enzyme defect, congenital anaemia
 CC such as sickle cell anaemia, Gaucher's disease, lysosomal storage
 CC diseases such as mucopolysaccharidosis, adrenal white matter
 CC degeneration, a variety of cancer and tumours. Proteins of the invention
 CC are useful for treating diseases such as Parkinson's disease, Alzheimer's
 CC disease and other neurodegenerative diseases, thrombocytopaenia, immune
 CC deficiencies and disorders such as severe combined immunodeficiency
 CC (SCID) and autoimmune disorders such as multiple sclerosis, systemic
 CC lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary
 CC inflammation. Sequences of the invention are also useful in gene therapy.
 CC The present sequence is stem cell growth factor-like protein from human
 XX Sequence 272 AA;
 SQ
 Query Match 99.3%; Score 1505; DB 4; Length 272;
 Best Local Similarity 100.0%; Pred. No. 2.8e-110;
 Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 HURLISWLFILNFMEYIGSQNASRRORRRHHPNVSQCGGCATCSDYNGCLSKPRL 62
 DB 2 HURLISWLFILNFMEYIGSQNASRRORRRHHPNVSQCGGCATCSDYNGCLSKPRL 61
 QY 63 PFALERIGMKQIGVCLSSPCSGYGYTRYPDINKTKCRADCDTCFNKNFCTKCKSGFYHL 122

Db 62 FFALERIGMKQIGVCLSSCPGYYGTRYPDINKCTCKADCTCFNNKFCCKSGFYLH 121
 Qy 123 LKCLDNCPEGLBANNHTMECVSIHVCEVSEWNPWSCTTKGKTCGPKRGTTETRVREIIQ 182
 Db 122 LKCLDNCPEGLBANNHTMECVSIHVCEVSEWNPWSCTTKGKTCGPKRGTTETRVREIIQ 181
 Qy 183 HPSAKGNLCPPTNTRKCTVQRKKCKGKRGKGRKRKKPKNGSKSAIPDSKSLESS 242
 Db 182 HPSAKGNLCPPTNTRKCTVQRKKCKGKRGKGRKRKKPKNGSKSAIPDSKSLESS 241
 Qy 243 KEIPEORENKQKKRKVKQDKQKSVSVTVH 273
 Db 242 KEIPEORENKQKKRKVKQDKQKSVSVTVH 272

RESULT 6

AAE13150
ID AAE13150 standard; protein; 272 AA.

AC AAE13150;

XX 28-JAN-2002 (first entry)

XX Human stem cell growth factor-like protein #2.

XX Human; stem cell growth factor-like protein; antiinflammatory; neutropic;
 KW neuroprotective; vulnerary; cytostatic; anticonvulsant; immunostimulant;
 KW vasotropic; virucide; dermatological; tranquilizer; cerebroprotective;
 KW osteopathic; immunodeficiency syndrome; chronic granulomatous disease;
 KW duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;
 KW acquired immune deficiency syndrome; agammaglobulinemia; thalassemia;
 KW Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;
 KW adrenal white matter degeneration; anaemia; neurodegenerative disease;
 KW Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;
 KW severe combined immunodeficiency; immune disorder; autoimmune disorder;
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
 KW autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;
 KW supporting factor for the proliferation of stem cell.

XX Homo sapiens.

XX Key Location/Qualifiers

FF Peptide 1..21

FT /label= signal_peptide

FT Protein 22..272

FT /note= "Human mature stem cell growth factor-like protein"

XX WO200177169-A2.

XX 18-OCT-2001.

XX 05-APR-2001; 2001WO-US011208.

XX 05-APR-2000; 2000US-00543774.

XX 28-JUN-2000; 2000US-0215733P.

XX 09-JAN-2001; 2000US-00757562.

XX 05-FEB-2001; 2001US-0266614P.

XX (HYSE-) HYSEO INC.

XX (KIRI) KIRIN BEER KK.

XX Tang TY, Labat I, Tillinghast JS, Sinku A, Liu C, Drmanac RT;

XX Stache-Crain B, Dickson M, Mize NK, Nishikawa M;

XX WPI; 2001-657166/75.

XX N-PSDB; AAD21724.

XX Novel stem cell growth factor like polypeptides and polynucleotides for
 PT identifying modulators useful for treating diseases such as Alzheimer's
 PT disease, cancer, rheumatoid arthritis, osteoporosis.

PS Claim 28; Page 211-212; 232pp; English.

XX The patent discloses novel stem cell growth factor-like proteins and
 CC polynucleotides encoding them. Proteins of the invention are also known
 CC as supporting factor for the proliferation of stem cells (SCR-1). Stem
 CC cell growth factor-like proteins are useful for supporting proliferation
 CC or survival of a stem cell or germ cell which is preferably primordial
 CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem
 CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent
 CC cell. The haematopoietic progenitor cell cultured using stem cell growth
 CC factor-like proteins can replace as a graft for the bone marrow
 CC transplantation or cord blood transplantation for treating a variety of
 CC diseases such as immunodeficiency syndrome, chronic granulomatous
 CC disease, duplicated immunodeficiency syndrome, agammaglobulinemia,
 CC Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),
 CC thalassemia, haemolytic anaemia due to enzyme defect, congenital anaemia
 CC such as sickle cell anaemia, Gaucher's disease, lysosomal storage
 CC diseases such as mucopolysaccharidosis, adrenal white matter
 CC degeneration, a variety of cancer and tumours. Proteins of the invention
 CC are useful for treating diseases such as Parkinson's disease, Alzheimer's
 CC disease and other neurodegenerative diseases, thrombocytopaenia, immune
 CC deficiencies and disorders such as severe combined immunodeficiency
 CC (SCID) and autoimmune disorders such as multiple sclerosis, systemic
 CC lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary
 CC inflammation. Sequences of the invention are also useful in gene therapy.
 CC The present sequence is stem cell growth factor-like protein from human

XX Sequence 272 AA;

Query Match 99.3%; Score 1505; DB 4; Length 272;

Best Local Similarity 100.0%; Pred. No. 2.8e-110;

Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 HLELISWLPILFMFMEYIGSNASRGRRORRMRHPNVSGCGGCATCDYNGCLSKPRL 62

Db 2 HLELISWLPILFMFMEYIGSNASRGRRORRMRHPNVSGCGGCATCDYNGCLSKPRL 61

Qy 63 FFALERIGMKQIGVCLSSCPGYYGTRYPDINKCTCKADCTCFNNKFCCKSGFYLH 122

Db 62 FFALERIGMKQIGVCLSSCPGYYGTRYPDINKCTCKADCTCFNNKFCCKSGFYLH 121

Qy 123 LKCLDNCPEGLBANNHTMECVSIHVCEVSEWNPWSCTTKGKTCGPKRGTTETRVREIIQ 182

Db 122 LKCLDNCPEGLBANNHTMECVSIHVCEVSEWNPWSCTTKGKTCGPKRGTTETRVREIIQ 181

Qy 183 HPSAKGNLCPPTNTRKCTVQRKKCKGKRGKGRKRKKPKNGSKSAIPDSKSLESS 242

Db 182 HPSAKGNLCPPTNTRKCTVQRKKCKGKRGKGRKRKKPKNGSKSAIPDSKSLESS 241

Qy 243 KEIPEORENKQKKRKVKQDKQKSVSVTVH 273

Db 242 KEIPEORENKQKKRKVKQDKQKSVSVTVH 272

RESULT 7

ABR62112

ID ABR62112 standard; protein; 272 AA.

XX ABR62112;

XX 18-AUG-2003 (first entry)

XX Human clone 1 thrombospondin protein #23.

XX Human; secreted; stem cell growth factor; cytostatic; haemostatic;
 KW neuroprotective; immunostimulant; leukaemia; haemophilia; cancer;
 KW degenerative disease; Alzheimer's disease; food supplement;
 KW immunological disorder; thrombospondin.

XX Homo sapiens.

XX WO2003029405-A2.

PD 10-APR-2003.
XX 30-AUG-2002; 2002WO-US027746.
XX 30-AUG-2001; 2001US-03163689.
PR 10-DEC-2001; 2001US-0339739P.
PR 19-APR-2002; 2002US-00125852.
XX (HYSE-) HYSEQ INC.
PA Tang YT;
PI WPI; 2003-381616/36.
DR New stem cell growth factor-like polypeptides and polynucleotides, useful
PT for treating e.g. leukemia, hemophilia and degenerative diseases like
PT Alzheimer's disease, and for inducing immune response.
XX Disclosure; Fig 1; 151pp; English.
XX The invention relates to new stem cell growth factor-like polypeptides
CC and polynucleotides. The stem cell growth factor-like polypeptides and
CC polynucleotides are useful for inducing differentiation of embryonic and
CC adult stem cells to give rise to different cell types, for treating e.g.
CC leukaemia, haemophilia and degenerative diseases like Alzheimer's
CC disease. They are also useful for generating new tissues and organs that
CC may aid patients in need of transplanted tissues. The polynucleotides are
CC useful as hybridisation probes, oligomers or primers for PCR, for
CC chromosome and gene mapping, in recombinantly producing protein, in
CC generating antisense DNA or RNA, in diagnostics as expressed sequence
CC tags for identifying expressed genes, and for inducing immune response.
CC The polypeptides are useful for generating antibodies that specifically
CC bind the polypeptide, as molecular weight markers, and as a food
CC supplement (e.g. protein or amino acid supplement, and as a carbon,
CC nitrogen or carbohydrate source). Compositions comprising the
CC polypeptides or polynucleotides are useful for the diagnosis, treatment
CC or prevention of cancers, and other immunological disorders. The current
CC sequence represents a human clone 1 thrombospondin protein
XX
SQ Sequence 272 AA;
Query Match 99.3%; Score 1505; DB 6; Length 272;
Best Local Similarity 100.0%; Pred. No. 2.8e-110;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 HURLISWLPILNFMFYIGSQNASRGRQRMRHPNVSGQGCGCATCSYNGCLCKPRL 62
Db |||||
2 HURLISWLPILNFMFYIGSQNASRGRQRMRHPNVSGQGCGCATCSYNGCLCKPRL 61
QY 63 PFALERIGMKQIGVCLSSCPGYYGTRYDINNKTKCKADCDTCFNKNFCTCKSGFYLL 122
Db |||||
62 PFALERIGMKQIGVCLSSCPGYYGTRYDINNKTKCKADCDTCFNKNFCTCKSGFYLL 121
QY 123 LGKCLDNCPEGLEANNHTECVSIVHCYSEWNPWSPCTKKGKTCGFRGTETRVREIIQ 182
Db |||||
122 LGKCLDNCPEGLEANNHTECVSIVHCYSEWNPWSPCTKKGKTCGFRGTETRVREIIQ 181
QY 183 HPSAKGNLCPTNETRKTQVRKKCKGKGRKKGRKKRKKPNKNGSKSAIPDSKLSLS 242
Db |||||
182 HPSAKGNLCPTNETRKTQVRKKCKGKGRKKGRKKRKKPNKNGSKSAIPDSKLSLS 241
QY 243 KEIPEORENKQKKKRVODKQKSVSVTVH 273
Db |||||
242 KEIPEORENKQKKKRVODKQKSVSVTVH 272
RESULT 8
ABO44432
ID ABO44432 standard; protein; 272 AA.
XX
AC ABO44432;
XX
DT 30-SEP-2003 (first entry)

XX Human stem cell growth factor-like protein, SCR 1 #4.
DE
XX
KW Stem cell growth factor-like protein; antianemic; anti-HIV; SCR-1;
KW immunostimulant; vulnery; haematopoietic stem cell; gene therapy;
KW supporting factor for proliferation of stem cells; wound healing;
KW haematopoietic progenitor cell; stromal cell; AIDS; thalassaemia;
KW bone marrow transplantation; cord blood transplantation;
KW chronic granulomatous disease; duplicated immunodeficiency syndrome;
KW agammaglobulinaemia; Wiskott-Aldrich syndrome; haemolytic anaemia;
KW congenital anaemia; sicklaemia; Gaucher's disease; morphogenesis;
KW epithelial cell growth; ovarian follicle development; nerve cell growth;
KW cartilage remodeling; bone growth; immunosuppression; human.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..21 /label= Signal_peptide
FT Protein 22..272 /note= "Mature stem cell growth factor-like protein. This
protein is specifically claimed in claim 9"
XX US2003044792-A1.
XX 06-MAR-2003.
XX 28-JUN-2001; 2001US-00894912.
XX 28-JUN-2000; 2000US-0215733P.
PR 05-FEB-2001; 2001US-0266614P.
PR 05-APR-2001; 2001US-0282397P.
XX (TANG/) TANG Y T.
PA (LABA/) LABAT I.
PA (DRMA/) DRMANAC R T.
PA (MIZE/) MIZE N.
PA (NISH/) NISHIKAWA M.
PA (CHAO/) CHAO C.
XX
PI Tang YT, Labat I, Drmanac RT, Mize N, Nishikawa M, Chao C;
XX WPI; 2003-625403/59.
DR N-PSDB; ACH04328.
XX
PT Novel isolated polypeptide having stem cell growth factor activity,
PT useful for promoting wound healing, and as a medicine to proliferate or
PT support human hematopoietic stem cells or human hematopoietic progenitor
PT cells.
XX
XX Claim 23; Page 82; 96pp; English.
XX The invention relates to an isolated stem cell growth factor-like
CC polypeptide (referred as supporting factor for proliferation of stem
CC cells (SCR-1)) from mouse or human, or its mature protein portion, or
CC fragment, analogue, variant or derivative, that retains stem cell growth
CC factor activity. Also included are an isolated polynucleotide encoding
CC SCR-1 (or its mature protein portion, or fragment, analogue, variant or
CC derivative, that retains stem cell growth factor activity, or the
CC complement of the polynucleotide), an (expression) vector comprising the
CC SCR-1 polynucleotide, a host cell genetically engineered to contain the
CC SCR-1 polynucleotide in operative association with a regulatory sequence
CC that controls expression of the polynucleotide in the host cell,
CC preparation of the SCR-1 polypeptide, a polypeptide which is an
CC expression product of the SCR-1 polynucleotide (the polypeptide having an
CC activity to support proliferation or survival of haematopoietic stem cell
CC or haematopoietic progenitor cell, with a proviso that C-terminal aa
CC sequence does not comprise the aa sequence appearing as ABO44433), an
CC isolated SCR-1 polypeptide with stem cell growth factor activity and
CC lacking any 10 consecutive aa from ABO44430, an isolated polypeptide
CC with stem cell growth factor activity having at least an aa sequence
CC appearing as ABO44428 and ABO44429, a culture medium comprising the SCR-1
CC polypeptide to maintain survival of or promote proliferation of a stem

CC cell or germ cell, an anti-SCR-1 antibody, a nucleic acid array
CC comprising the SCR-1 polynucleotide or a unique segment of the SCR-1
CC polynucleotide attached to a surface, a stromal cell genetically
CC engineered to express the SCR-1 polypeptide to support proliferation or
CC survival of a stem cell or germ cell and an implant comprising a cell
CC genetically engineered to express the SCR-1 polypeptide to support
CC proliferation or survival of a stem cell or germ cell. The SCR-1
CC polypeptide is useful for identifying a compound that binds to the SCR-1
CC polypeptide and for maintaining survival of or promoting proliferation of
CC a stem cell, a germ cell, a haematopoietic stem cell or a haematopoietic
CC progenitor cell. The SCR-1 polypeptide is useful for promoting wound
CC healing. The human haematopoietic stem cell or human haematopoietic
CC progenitor cell culture using the SCR-1 polypeptide can replace as a
CC graft for the conventional bone marrow transplantation or cord blood
CC transplantation. The transplantation of haematopoietic stem cells can be
CC employed as a therapy for treating diseases such as chronic granulomatous
CC diseases, duplicated immunodeficiency syndrome, agammaglobulinaemia,
CC Wiskott-Aldrich syndrome, AIDS, etc., thalassaemia, haemolytic anaemia
CC due to enzyme defect, congenital anaemia such as sickle cell, Gaucher's
CC disease etc. the SCR-1 polypeptide is useful for cell growth and
CC morphogenesis, including tissue specific stem cell growth, epithelial
CC cell growth and regulation, ovarian follicle development, promoting nerve
CC cell growth, sustaining neuronal populations, cartilage remodeling, bone
CC growth and immunosuppression. The present sequence is a Human SCR-1
CC protein
XX
SQ Sequence 272 AA;

Query Match 99.3%; Score 1505; DB 6; Length 272;

Best Local Similarity 100.0%; Pred. No. 2.8e-110;

Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HRLISLWFLILMFYIGSNASRGRRQRMRHPNVISQGGCATCSDYNGCLSKPRL 62
DB 2 HRLISLWFLILMFYIGSNASRGRRQRMRHPNVISQGGCATCSDYNGCLSKPRL 61
QY 63 FPALEIRIGMKQIGVCLSSCPGYYGTRYPDINKCTKACDCTCFNKNFTCKKSGPYLH 122
DB 62 FPALEIRIGMKQIGVCLSSCPGYYGTRYPDINKCTKACDCTCFNKNFTCKKSGPYLH 121
QY 123 LGKCLDNCPEGLBANNHTMVCVSIHVCEVSEWNPSPCTKKGKTCGPKRGTETRVREIIQ 162
DB 122 LGKCLDNCPEGLBANNHTMVCVSIHVCEVSEWNPSPCTKKGKTCGPKRGTETRVREIIQ 161
QY 183 HPSAKGNLCPPTNTRKCTVQRKCKGKGRKGRKRPKNKGSKEAIPDSKLSLESS 242
DB 182 HPSAKGNLCPPTNTRKCTVQRKCKGKGRKGRKRPKNKGSKEAIPDSKLSLESS 241
QY 243 KEIPEORENKQKKRQVDPKQKSVSVSTVH 273
DB 242 KEIPEORENKQKKRQVDPKQKSVSVSTVH 272

RESULT 9

ABO44413

ID ABO44413 standard; protein; 272 AA.

AC ABO44413;

XX ABO44413;

DT 30-SEP-2003 (first entry)

XX Human stem cell growth factor-like protein, SCR 1 #1.

XX Stem cell growth factor-like protein; antianemic; anti-HIV; SCR-1;
KW immunostimulant; vulnary; haematopoietic stem cell; gene therapy;
KW supporting factor for proliferation of stem cells; wound healing;
KW haematopoietic progenitor cell; stromal cell; AIDS; thalassaemia;
KW bone marrow transplantation; cord blood transplantation;
KW chronic granulomatous disease; duplicated immunodeficiency syndrome;
KW agammaglobulinaemia; Wiskott-Aldrich syndrome; haemolytic anaemia;
KW congenital anaemia; sickle cell; Gaucher's disease; morphogenesis;
KW epithelial cell growth; ovarian follicle development; nerve cell growth;
KW cartilage remodeling; bone growth; immunosuppression; human.

XX Homo sapiens.
OS
XX US2003044792-A1.
XX
XX PD 06-MAR-2003.
XX
XX PF 28-JUN-2001; 2001US-00894912.
XX
XX PR 28-JUN-2000; 2000US-0215733P.
XX
XX PR 05-FEB-2001; 2001US-0266614P.
XX
XX PR 05-APR-2001; 2001US-0282397P.
XX
XX (TANG/) TANG Y T.
XX (LABA/) LABAT I.
XX (DRMA/) DRMANAC R T.
XX (MIZE/) MIZE N.
XX (NISH/) NISHIKAWA M.
XX (CHAO/) CHAO C.
XX
XX PT Tang YT, Labat I, Drmanac RT, Mize N, Nishikawa M, Chao C;
XX
XX DR WPI; 2003-625403/59.
XX
XX DR N-PSDB; ACH04323, ACH04324.
XX
XX
XX PS Claim 23; Fig 3; 96pp; English.
XX
XX
XX CC The invention relates to an isolated stem cell growth factor-like
XX polypeptide (referred as supporting factor for proliferation of stem
XX cells (SCR-1)) from mouse or human, or its mature protein portion, or
XX fragment, analogue, variant or derivative, that retains stem cell growth
XX factor activity. Also included are an isolated polynucleotide encoding
XX SCR-1 or its mature protein portion, or fragment, analogue, variant or
XX derivative, that retains stem cell growth factor activity, or the
XX complement of the polynucleotide, an (expression) vector comprising the
XX SCR-1 polynucleotide, a host cell genetically engineered to contain the
XX SCR-1 polynucleotide in operative association with a regulatory sequence
XX that controls expression of the polynucleotide in the host cell,
XX preparation of the SCR-1 polypeptide, a polypeptide which is an
XX expression product of the SCR-1 polynucleotide (the polypeptide having an
XX activity to support proliferation or survival of haematopoietic stem cell
XX or haematopoietic progenitor cell, with a proviso that C-terminal aa
XX sequence does not comprise the aa sequence appearing as ABO44433), an
XX isolated SCR-1 polypeptide with stem cell growth factor activity and
XX lacking any 10 consecutive aas from ABO44430, an isolated polypeptide
XX with stem cell growth factor activity having at least an aa sequence
XX appearing as ABO44428 and ABO44429, a culture medium comprising the SCR-1
XX polypeptide to maintain survival of or promote proliferation of a stem
XX cell or germ cell, an anti-SCR-1 antibody, a nucleic acid array
XX comprising the SCR-1 polynucleotide or a unique segment of the SCR-1
XX polynucleotide attached to a surface, a stromal cell genetically
XX engineered to express the SCR-1 polypeptide to support proliferation or
XX survival of a stem cell or germ cell and an implant comprising a cell
XX genetically engineered to express the SCR-1 polypeptide to support
XX proliferation or survival of a stem cell or germ cell. The SCR-1
XX polypeptide is useful for identifying a compound that binds to the SCR-1
XX polypeptide and for maintaining survival of or promoting proliferation of
XX a stem cell, a germ cell, a haematopoietic stem cell or a haematopoietic
XX progenitor cell. The SCR-1 polypeptide is useful for promoting wound
XX healing. The human haematopoietic stem cell or human haematopoietic
XX progenitor cell culture using the SCR-1 polypeptide can replace as a
XX graft for the conventional bone marrow transplantation or cord blood
XX transplantation. The transplantation of haematopoietic stem cells can be
XX employed as a therapy for treating diseases such as chronic granulomatous
XX diseases, duplicated immunodeficiency syndrome, agammaglobulinaemia,
XX Wiskott-Aldrich syndrome, AIDS, etc., thalassaemia, haemolytic anaemia
XX due to enzyme defect, congenital anaemia such as sickle cell, Gaucher's
XX disease etc. the SCR-1 polypeptide is useful for cell growth and

CC morphogenesis, including tissue specific stem cell growth, epithelial
 CC cell growth and regulation, ovarian follicle development, promoting nerve
 CC cell growth, sustaining neuronal populations, cartilage remodeling, bone
 CC growth and immunosuppression. The present sequence is a Human SCR-1
 CC protein
 XX
 SQ Sequence 272 AA;

Query Match 99.3%; Score 1505; DB 6; Length 272;
 Best Local Similarity 100.0%; Pred. No. 2.8e-110;
 Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HRLISWLFILNMFYIGSONASRGRRRRMHPNVISQGGCATCDYNGCLSKPRL 62
 DB 2 HRLISWLFILNMFYIGSONASRGRRRRMHPNVISQGGCATCDYNGCLSKPRL 61
 QY 63 PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKCTCKADCDTCFNKNFCTCKKSGPYLH 122
 DB 62 PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKCTCKADCDTCFNKNFCTCKKSGPYLH 121
 QY 123 LGKCLDNCPEGLANNNHTECVSIHVCEVSEWNPSPCTTKGKTCGPKGTETVRVRIIQ 182
 DB 122 LGKCLDNCPEGLANNNHTECVSIHVCEVSEWNPSPCTTKGKTCGPKGTETVRVRIIQ 181
 QY 183 HPSAKGNLCPPTNETRKTCTVQRKCKQGERGKGRERKRPKNKESKEAIPDSKSLSS 242
 DB 182 HPSAKGNLCPPTNETRKTCTVQRKCKQGERGKGRERKRPKNKESKEAIPDSKSLSS 241
 QY 243 KEIPEQRENKQOQKRVQDKQKS 273
 DB 242 KEIPEQRENKQOQKRVQDKQKS 272

RESULT 10
 ID AAE13163
 AC AAE13163;
 XX
 XX
 DT 28-JAN-2002 (first entry)
 XX
 XX Human secreted protein from clone DA228_6.

XX Human; stem cell growth factor-like protein; antiinflammatory; neurotropic;
 KW neuroprotective; vulnerary; cytostatic; anticonvulsant; immunostimulant;
 KW vasotropic; virucide; dermatological; tranquiliser; cerebroprotective;
 KW osteopathic; immunodeficiency syndrome; chronic granulomatous disease;
 KW duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;
 KW acquired immune deficiency syndrome; agammaglobulinaemia; thalassaemia;
 KW Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;
 KW adrenal white matter degeneration; anaemia; neurodegenerative disease;
 KW Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;
 KW severe combined immunodeficiency; immune disorder; autoimmune disorder;
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
 KW autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;
 KW supporting factor for the proliferation of stem cell; secreted protein.

XX Homo sapiens.

OS
 XX
 PN WD200177169-A2.
 XX
 XX 18-OCT-2001.

PF 05-APR-2001; 2001MO-US011208.

XX 05-APR-2000; 2000US-00543774.

PR 28-JUN-2000; 2000US-0215733P.

PR 09-JAN-2001; 2001US-00752562.

PR 05-FEB-2001; 2001US-0266614P.

XX (HYSE-) HYSEQ INC.

PA (KIRI) KIRIN BEER KK.

XX

PI

PI

XX

XX

XX

XX

PT

PT

PT

XX

XX

XX

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

Sequence 265 AA;

Query Match

Best Local Similarity

Matches 264;

Conservative 0;

Mismatches 0;

Indels 0;

Gaps 0;

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

RESULT 11

ABO44427

ID ABO44427

XX ABO44427

XX ABO44427;

XX 30-SEP-2003 (first entry)

XX Human secreted protein clone DA228_6.

XX

XX

Tang TY, Labat I, Tillinghast JS, Sinku A, Liu C, Drmanac RT;
 Stache-Crain B, Dickson M, Mize NK, Nishikawa M;
 WPI; 2001-657166/75.

Novel stem cell growth factor like polypeptides and polynucleotides for
 identifying modulators useful for treating diseases such as Alzheimer's
 disease, cancer, rheumatoid arthritis, osteoporosis.

Disclosure; Fig 3; 232pp; English.

The patent discloses novel stem cell growth factor-like proteins and
 polynucleotides encoding them. Proteins of the invention are also known
 as supporting factor for the proliferation of stem cells (SCR-1). Stem
 cell growth factor-like proteins are useful for supporting proliferation
 or survival of a stem cell or germ cell which is preferably primordial
 germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem
 cell, haematopoietic progenitor cell, pluripotent cell or totipotent
 cell. The haematopoietic progenitor cell cultured using stem cell growth
 factor-like proteins can replace as a graft for the bone marrow
 transplantation or cord blood transplantation for treating a variety of
 diseases such as immunodeficiency syndrome, chronic granulomatous
 disease, duplicated immunodeficiency syndrome, agammaglobulinaemia,
 Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),
 thalassaemia, haemolytic anaemia due to enzyme defect, congenital anaemia
 such as sickle cell anaemia, Gaucher's disease, lysosomal storage
 diseases such as mucopolysaccharidosis, adrenal white matter
 degeneration, a variety of cancer and tumours. Proteins of the invention
 are useful for treating diseases such as Parkinson's disease, Alzheimer's
 disease and other neurodegenerative diseases, thrombocytopaenia, immune
 deficiencies and disorders such as severe combined immunodeficiency
 (SCID) and autoimmune disorders such as multiple sclerosis, systemic
 lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary
 inflammation. Sequences of the invention are also useful in gene therapy.
 The present sequence is human secreted protein from clone DA228_6

Query Match

Best Local Similarity

Matches 264;

Conservative 0;

Mismatches 0;

Indels 0;

Gaps 0;

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

Stem cell growth factor-like protein; antianemic; anti-HIV; SCR-1; immunostimulant; vulnary; haematopoietic stem cell; gene therapy; supporting factor for proliferation of stem cells; wound healing; haematopoietic progenitor cell; stromal cell; AIDS; thalassaemia; bone marrow transplantation; cord blood transplantation; chronic granulomatous disease; duplicated immunodeficiency syndrome; agammaglobulinaemia; Wiskott-Aldrich syndrome; haemolytic anaemia; congenital anaemia; sicklaemia; Gaucher's disease; morphogenesis; epithelial cell growth; ovarian follicle development; nerve cell growth; cartilage remodeling; bone growth; immunosuppression; human.

OS Homo sapiens.

XX US2003044792-A1.

XX 06-MAR-2003.

XX 28-JUN-2001; 2001US-00894912.

XX 28-JUN-2000; 2000US-0215733P.

XX 05-FEB-2001; 2001US-0266614P.

XX 05-APR-2001; 2001US-0282397P.

XX (TANG/) TANG Y T.

XX (LABA/) LABAT I.

XX (DRMA/) DRMANAC R T.

XX (MIZE/) MIZE N.

XX (NISH/) NISHIKAWA M.

XX (CHAO/) CHAO C.

XX Tang YT, Labat I, Drmanac RT, Mize N, Nishikawa M, Chao C;

XX WPI; 2003-625403/59.

XX Novel isolated polypeptide having stem cell growth factor activity.

XX useful for promoting wound healing, and as a medicine to proliferate or

XX support human hematopoietic stem cells or human hematopoietic progenitor

XX cells.

XX Disclosure; Fig 3; 96pp; English.

XX The invention relates to an isolated stem cell growth factor-like

XX polypeptide (referred as supporting factor for proliferation of stem

XX cells (SCR-1)) from mouse or human, or its mature protein portion, or

XX fragment, analogue, variant or derivative, that retains stem cell growth

XX factor activity. Also included are an isolated polynucleotide encoding

XX SCR-1 (or its mature protein portion, or fragment, analogue, variant or

XX derivative, that retains stem cell growth factor activity, or the

XX complement of the polynucleotide), an (expression) vector comprising the

XX SCR-1 polynucleotide, a host cell genetically engineered to contain the

XX SCR-1 polynucleotide in operative association with a regulatory sequence

XX that controls expression of the polynucleotide in the host cell,

XX preparation of the SCR-1 polypeptide, a polypeptide which is an

XX expression product of the SCR-1 polynucleotide (the polypeptide having an

XX activity to support proliferation or survival of haematopoietic stem cell

XX or haematopoietic progenitor cell, with a proviso that C-terminal aa

XX sequence does not comprise the aa sequence appearing as ABO44433), an

XX isolated SCR-1 polypeptide with stem cell growth factor activity and

XX lacking any 10 consecutive aas from ABO44430, an isolated polypeptide

XX with stem cell growth factor activity having at least an aa sequence

XX appearing as ABO44428 and ABO44429, a culture medium comprising the SCR-1

CC healing. The human haematopoietic stem cell or human haematopoietic
 CC progenitor cell culture using the SCR-1 polypeptide can replace as a
 CC graft for the conventional bone marrow transplantation or cord blood
 CC transplantation. The transplantation of haematopoietic stem cells can be
 CC employed as a therapy for treating diseases such as chronic granulomatous
 CC diseases, duplicated immunodeficiency syndrome, agammaglobulinaemia,
 CC Wiskott-Aldrich syndrome, AIDS, etc., thalassaemia, haemolytic anaemia
 CC due to enzyme defect, congenital anaemia such as sicklaemia, Gaucher's
 CC disease etc. the SCR-1 polypeptide is useful for cell growth and
 CC morphogenesis, including tissue specific stem cell growth, epithelial
 CC cell growth and regulation, ovarian follicle development, promoting nerve
 CC cell growth, sustaining neuronal populations, cartilage remodeling, bone
 CC growth and immunosuppression. The present sequence is a protein
 CC homologous to Human SCR-1 protein
 XX

SQ Sequence 265 AA;

Query Match 97.1%; Score 1472; DB 6; Length 265;

Best Local Similarity 100.0%; Pred. No. 1.1e-107; Mismatches 0; Gaps 0;

Matches 264; Conservative 0; Indels 0;

Qy 3 HLRLLSWLFIILNFMVEYIGSQNASRRQRHHPNVSGCGGCATCSDYNGCLSKPRL 62

Db 2 HLRLLSWLFIILNFMVEYIGSQNASRRQRHHPNVSGCGGCATCSDYNGCLSKPRL 61

Qy 63 PFALERIGMKOIGVCLSSCPGGYTGTRYPDINKTKKADCDTCFNKQFCTKCKSGFYIH 122

Db 62 PFALERIGMKOIGVCLSSCPGGYTGTRYPDINKTKKADCDTCFNKQFCTKCKSGFYIH 121

Qy 123 LGKCLDNCPEGLRANNHTMECVSIHVCEVSWNPWSPCTKKGKTCGPKRGTRVREIIQ 182

Db 122 LGKCLDNCPEGLRANNHTMECVSIHVCEVSWNPWSPCTKKGKTCGPKRGTRVREIIQ 181

Qy 183 HPSAKGNLCPTNETRKTCTVQRKCKGGERGKGRERKRPKNKGESKEAIPDSKSLESS 242

Db 182 HPSAKGNLCPTNETRKTCTVQRKCKGGERGKGRERKRPKNKGESKEAIPDSKSLESS 241

Qy 243 KEIPEORENKQOQKRRKVDQKQS 266

Db 242 KEIPEORENKQOQKRRKVDQKQS 265

RESULT 12

AAW85607

ID AAW85607 standard; protein; 292 AA.

XX AAW85607;

XX 02-MAR-1999 (first entry)

XX Secreted protein clone da228_6.

XX Clone; secreted protein; protein factor; cytokine; lymphokine;

XX interferon; colony stimulating factor; CSF; interleukin; cloning;

XX tumour invasion; tumour suppression; immune boosting.

XX Homo sapiens.

XX WO9849302-A1.

XX 05-NOV-1998.

XX 24-APR-1998; 98WO-US008336.

XX 25-APR-1997; 97US-00845296.

XX 23-APR-1998; 98US-00065125.

XX (GENY) GENETICS INST INC.

XX Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;

XX Spaulding V, Agostino MJ;

XX WPI; 1999-024059/02.

DR N-PSDB; AAV83133.
XX New polynucleotides encoding secreted human proteins - are derived from
PT human foetal brain, adult brain, adult blood or placenta cDNA libraries,
PT useful, e.g. as potential immunomodulators.
PS Claim 8; Page 63-64; 104pp; English.
XX The nucleotide sequence (NS) of the full-length protein-coding sequence
CC of clones c1254 (AAV83132), da2286 (AAV83133), dh4105 (AAV83134), eh801
CC (AAV83135), er3691 (AAV83136), fh1235 (AAV83137), fm601 (AAV83138) or
CC fr4732 (AAV83139). (all clones are deposited as ATCC 98415) and the
CC proteins they encode are predicted to have biological activities which
CC would make them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals for example, tumour suppression/invasion
CC activity, immune system boosting activity. The polynucleotides are also
CC believed to be useful for gene therapy
XX Sequence 292 AA;
XX Query Match 97.1%; Score 1472; DB 2; Length 292;
XX Best Local Similarity 100.0%; Pred. No. 1.2e-107;
XX Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 HLLRLISWLFILNFMFYIGSQNASRGRORRMRHPNVSQCGGCATCSDYNGCLSCKPRLL 62
DB 2 HLLRLISWLFILNFMFYIGSQNASRGRORRMRHPNVSQCGGCATCSDYNGCLSCKPRLL 61
QY 63 FFALEIRIGMKQIGVCLSSCPGSGYGYTRYPDINKCTKACDCTCFNKNFTCKSGFYLLH 122
DB 62 FFALEIRIGMKQIGVCLSSCPGSGYGYTRYPDINKCTKACDCTCFNKNFTCKSGFYLLH 121
QY 123 LGKCLDNCPEGLBANNHTMCEVSIHVCEVSEWNPSPCTKKGTKCGFRGTETVRVRIIQ 182
DB 122 LGKCLDNCPEGLBANNHTMCEVSIHVCEVSEWNPSPCTKKGTKCGFRGTETVRVRIIQ 181
QY 183 HPSAKGNLCPPTNETRKTCTVQRKCKGGRKGRKRKPKNKGSKRAIPDSKLESS 242
DB 182 HPSAKGNLCPPTNETRKTCTVQRKCKGGRKGRKRKPKNKGSKRAIPDSKLESS 241
QY 243 KEIPEQRENKQKQKRVQDKQS 266
DB 242 KEIPEQRENKQKQKRVQDKQS 265
RESULT 13
ID AAE13170 standard; protein; 292 AA.
XX AAE13170;
XX AC AAE13170;
XX DT 28-JAN-2002 (first entry)
XX DE Human SCR-1 related protein.
XX KW Human; stem cell growth factor-like protein; antiinflammatory; neurotropic;
KW neuroprotective; vulnerary; cytostatic; anticonvulsant; immunostimulant;
KW vasotropic; virucide; dermatological; tranquiliser; cerebroprotective;
KW osteopathic; immunodeficiency syndrome; chronic granulomatous disease;
KW duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;
KW acquired immune deficiency syndrome; agammaglobulinaemia; thalassaemia;
KW Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;
KW adrenal white matter degeneration; anaemia; neurodegenerative disease;
KW Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;
KW severe combined immunodeficiency; immune disorder; autoimmune disease;
KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
KW autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;
KW supporting factor for the proliferation of stem cell.
XX Unidentified.
XX OS WO200177169-A2.
XX PN
XX

PD 18-OCT-2001.
XX 05-APR-2001; 2001WO-US011208.
XX 05-APR-2000; 2000US-00543774.
XX 28-JUN-2000; 2000US-0215733P.
XX 09-JAN-2001; 2001US-00757562.
XX 05-FEB-2001; 2001US-0266614P.
XX (HYSB-) HYSQ INC.
XX (KIRI) KIRIN BEER KK.
XX Tang TY, Labat I, Tillinghast JS, Sinku A, Liu C, Drmanac RT;
XX Stache-Crain B, Dickson M, Mize NK, Nishikawa M;
XX WPI; 2001-657166/75.
XX N-PSDB; AAD21740.
XX Novel stem cell growth factor like polypeptides and polynucleotides for
XX identifying modulators useful for treating diseases such as Alzheimer's
XX disease, cancer, rheumatoid arthritis, osteoporosis.
XX Claim 27; Page 231-232; 232pp; English.
XX The patent discloses novel stem cell growth factor-like proteins and
XX polynucleotides encoding them. Proteins of the invention are also known
XX as supporting factor for the proliferation of stem cells (SCR-1). Stem
XX cell growth factor-like proteins are useful for supporting proliferation
XX or survival of a stem cell or germ cell which is preferably primordial
XX germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem
XX cell, haematopoietic progenitor cell, pluripotent cell or totipotent
XX cell. The haematopoietic progenitor cell cultured using stem cell growth
XX factor-like proteins can replace as a graft for the bone marrow
XX transplantation or cord blood transplantation for treating a variety of
XX diseases such as immunodeficiency syndrome, chronic granulomatous
XX disease, duplicated immunodeficiency syndrome, agammaglobulinaemia,
XX Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),
XX thalassaemia, haemolytic anaemia due to enzyme defect, congenital anaemia
XX such as sickle cell anaemia, Gaucher's disease, lysosomal storage
XX diseases such as mucopolysaccharidosis, adrenal white matter
XX degeneration, a variety of cancer and tumours. Proteins of the invention
XX are useful for treating diseases such as Parkinson's disease, Alzheimer's
XX disease and other neurodegenerative diseases, thrombocytopaenia, immune
XX deficiencies and disorders such as severe combined immunodeficiency
XX (SCID) and autoimmune disorders such as multiple sclerosis, systemic
XX lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary
XX inflammation. Sequences of the invention are also useful in gene therapy.
XX The present sequence is human SCR-1 related protein
XX Sequence 292 AA;
XX Query Match 97.1%; Score 1472; DB 4; Length 292;
XX Best Local Similarity 100.0%; Pred. No. 1.2e-107;
XX Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 HLLRLISWLFILNFMFYIGSQNASRGRORRMRHPNVSQCGGCATCSDYNGCLSCKPRLL 62
DB 2 HLLRLISWLFILNFMFYIGSQNASRGRORRMRHPNVSQCGGCATCSDYNGCLSCKPRLL 61
QY 63 FFALEIRIGMKQIGVCLSSCPGSGYGYTRYPDINKCTKACDCTCFNKNFTCKSGFYLLH 122
DB 62 FFALEIRIGMKQIGVCLSSCPGSGYGYTRYPDINKCTKACDCTCFNKNFTCKSGFYLLH 121
QY 123 LGKCLDNCPEGLBANNHTMCEVSIHVCEVSEWNPSPCTKKGTKCGFRGTETVRVRIIQ 182
DB 122 LGKCLDNCPEGLBANNHTMCEVSIHVCEVSEWNPSPCTKKGTKCGFRGTETVRVRIIQ 181
QY 183 HPSAKGNLCPPTNETRKTCTVQRKCKGGRKGRKRKPKNKGSKRAIPDSKLESS 242
DB 182 HPSAKGNLCPPTNETRKTCTVQRKCKGGRKGRKRKPKNKGSKRAIPDSKLESS 241
QY 243 KEIPEQRENKQKQKRVQDKQS 266
XX

Db 242 KEIPEQRENKQQKKRKVKQDKQS 265

RESULT 14
ABP61846
ID ABP61846 standard; protein; 292 AA.
AC ABP61846;
XX
DT 04-OCT-2002 (first entry)
XX
DE Human polypeptide SEQ ID NO 200.
XX
KW Human, cytostatic; antirheumatic; antiarthritic; vulnary; analgesic;
KW antiinflammatory; antibacterial; immunosuppressive; antiparkinsonian;
KW neuroprotective; nootropic; osteopathic; haemostatic; vasotrophic;
KW antitumor; fungicide; antidiabetic; antiaesthetic; antiallergic;
KW immunostimulant; antiparasitic; secreted protein; transmembrane protein;
KW cytokine; cell proliferation; cell differentiation; autoimmune disease;
KW stem cell; growth factor; nervous system disease; neuropathy;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW osteoporosis; severe combined immunodeficiency; SCID; infection;
KW multiple sclerosis; rheumatoid arthritis; gene therapy.
XX
OS Homo sapiens.
XX
PN US2002065394-A1.
XX
PD 30-MAY-2002.
XX
PP 22-DEC-2000; 2000US-00745763.
XX
PR 18-MAR-1998; 98US-00040963.
XX
PA (JACO/) JACOBS K.
PA (MCCO/) MCCOY J M.
PA (LAVA/) LAVALLIE E R.
PA (COLL/) COLLINS-RACIE L A.
PA (EVAN/) EVANS C.
PA (MERB/) MERBERG D.
PA (TREA/) TREACY M.
PA (SPAU/) SPAULDING V.
XX
PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Spaulding V;
XX
DR WPI; 2002-582343/62.
DR N-PSDB; ABQ92060.
XX
PT Novel secreted or transmembrane protein and polynucleotide encoding the
PT protein, useful for diagnosis and treatment of neurological disorders,
PT cancer, autoimmune diseases, bone disorders and lung or liver fibrosis.
XX
PS Claim 207; Page 203-204; 284pp; English.
XX
CC The invention relates to human secreted or transmembrane protein (I),
CC their fragments and is encoded by specific complementary deoxyribonucleic
CC acid (cDNA) inserts (II), where the protein is substantially free from
CC other mammalian proteins. (I) are useful for preventing, treating or
CC ameliorating a medical condition, especially immunological treatment or
CC prevention of tumours. (I) exhibits activity relating to angiogenesis,
CC cytokine, cell proliferation, cell differentiation, antiinflammatory,
CC stem cell growth factor activity and activin or inhibin-related
CC activities. (I) can be used to manipulate stem cells in culture to give
CC rise to neuroepithelial cells that can be used to augment or replace
CC cells damaged by illness, autoimmune disease, accidental damage or
CC genetic disorders. (I) induces the proliferation of neural cells and
CC regeneration of nerve and brain tissue and is useful for the treatment of
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis. (I) is involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis and is useful for treating myeloid
CC or lymphoid cell disorders, platelet disorders such as thrombocytopaenia

CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve
CC tissue growth and in tissue repair, healing of burns, incisions, ulcers, or
CC for treating osteoporosis, osteoarthritis, bone degenerative disorders or
CC periodontal disease. (I) is also useful for gut protection or
CC regeneration and treatment of lung or liver fibrosis, reperfusion injury
CC in various tissues, various immune deficiencies and disorders including
CC severe combined immunodeficiency (SCID), bacterial or fungal infections,
CC autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis,
CC diabetes mellitus, myasthenia gravis, allergic reactions and conditions,
CC such as asthma or other respiratory problems. (II) is useful to express
CC recombinant protein, as markers for tissues in which the corresponding
CC protein is preferentially expressed and in gene therapy. The present
CC sequence is that of a polypeptide of the invention
XX
SQ Sequence 292 AA;
Query Match 97.1%; Score 1472; DB 5; Length 292;
Best Local Similarity 100.0%; Pred. No. 1.2e-107;
Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 HLRILSHLPILNPMVEYIGSNASGRGRORRHHNPNVSGCGGCGCATCSDYNGCLSCXPRL 62
Db 2 HLRILSHLPILNPMVEYIGSNASGRGRORRHHNPNVSGCGGCGCATCSDYNGCLSCXPRL 61
Qy 63 PFALERIGMKQIGVCLSSCPSSGVYGYTRYPDINKCTKCKADCCTCFNKNFCTKCKSGFYLN 122
Db 62 PFALERIGMKQIGVCLSSCPSSGVYGYTRYPDINKCTKCKADCCTCFNKNFCTKCKSGFYLN 121
Qy 123 LGKCLDNCPEGLEANNHTMBCVSIHVCEVSEWNPWSPCTKKGKTCGFKRGTRVREIIQ 182
Db 122 LGKCLDNCPEGLEANNHTMBCVSIHVCEVSEWNPWSPCTKKGKTCGFKRGTRVREIIQ 181
Qy 183 HPSAKGNLCPTTNETRKTCTVORKKCKGGRGKGRERKRKPKNGESKEAIPDSKSLSS 242
Db 182 HPSAKGNLCPTTNETRKTCTVORKKCKGGRGKGRERKRKPKNGESKEAIPDSKSLSS 241
Qy 243 KEIPEQRENKQQKKRKVKQDKQS 266
Db 242 KEIPEQRENKQQKKRKVKQDKQS 265
RESULT 15
ABR62114
ID ABR62114 standard; protein; 292 AA.
XX
AC ABR62114;
XX
DT 18-AUG-2003 (first entry)
XX
DE Human secreted protein clone da_288_6 #25.
XX
KW Human; secreted; stem cell growth factor; cytostatic; haemostatic;
KW neuroprotective; immunostimulant; leukaemia; haemophilia; cancer;
KW degenerative disease; Alzheimer's disease; food supplement;
KW immunological disorder.
XX
OS Homo sapiens.
XX
PN WO2003029405-A2.
XX
PD 10-APR-2003.
XX
PF 30-AUG-2002; 2002WO-US027746.
XX
PR 30-AUG-2001; 2001US-0316368P.
PR 10-DEC-2001; 2001US-0339739P.
PR 19-APR-2002; 2002US-00125852.
XX
PA (HYSB-) HYSBQ INC.
XX
PI Tang YT;
XX
DR WPI; 2003-381616/36.

XX New stem cell growth factor-like polypeptides and polynucleotides, useful
PT for treating e.g. leukemia, hemophilia and degenerative diseases like
PT Alzheimer's disease, and for inducing immune response.
XX
PS Disclosure; Fig 2; 151pp; English.
XX
XX The invention relates to new stem cell growth factor-like polypeptides
CC and polynucleotides. The stem cell growth factor-like polypeptides and
CC polynucleotides are useful for inducing differentiation of embryonic and
CC adult stem cells to give rise to different cell types, for treating e.g.
CC leukaemia, haemophilia and degenerative diseases like Alzheimer's
CC disease. They are also useful for generating new tissues and organs that
CC may aid patients in need of transplanted tissues. The polynucleotides are
CC useful as hybridisation probes, oligomers or primers for PCR, for
CC chromosome and gene mapping, in recombinantly producing protein, in
CC generating antisense DNA or RNA, in diagnostics as expressed sequence
CC tags for identifying expressed genes, and for inducing immune response.
CC The polypeptides are useful for generating antibodies that specifically
CC bind the polypeptide, as molecular weight markers, and as a food
CC supplement (e.g. protein or amino acid supplement, and as a carbon,
CC nitrogen or carbohydrate source). Compositions comprising the
CC polypeptides or polynucleotides are useful for the diagnosis, treatment
CC or prevention of cancers, and other immunological disorders. The current
CC sequence represents a human secreted protein clone da_288_6
XX
SQ Sequence 292 AA;

Query Match 97.1%; Score 1472; DB 6; Length 292;
Best Local Similarity 100.0%; Pred. No. 1.2e-107;
Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HLELIISWLFIIINMEYIGSNASRGRRORRHPNVSQCGGCATCSYNGCLCKPRL 62
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||
2 HLELIISWLFIIINMEYIGSNASRGRRORRHPNVSQCGGCATCSYNGCLCKPRL 61
QY 63 PFALERIGMKQIGVCLSSCPSPGYGYTRYPDINKCTKADCDTCFNKNFCTCKSGFYILH 122
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||
62 PFALERIGMKQIGVCLSSCPSPGYGYTRYPDINKCTKADCDTCFNKNFCTCKSGFYILH 121
QY 123 LGKCLDNCPGLEANNHTECVSIVHCEVSENNPWSFCTKKGKTCGTETRVREIIQ 182
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||
122 LGKCLDNCPGLEANNHTECVSIVHCEVSENNPWSFCTKKGKTCGTETRVREIIQ 181
QY 183 HPSAKGNLCPTNETRKCTVQRKCKGKGRKRRKKNKGSKEAIPDSKSLSS 242
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||
182 HPSAKGNLCPTNETRKCTVQRKCKGKGRKRRKKNKGSKEAIPDSKSLSS 241
QY 243 KEIPEORENKQOKRKRKVDOKS 266
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||
242 KEIPEORENKQOKRKRKVDOKS 265

Search completed: June 29, 2004, 16:59:51
Job time : 66.6986 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 16:57:05 ; Search time 14.2464 Seconds
(without alignments)
989.298 Million cell updates/sec

Title: US-09-894-912A-13
Perfect score: 1516
Sequence: 1 MGHRLISMLFIILNFWYI.....QQKKRKVDKQKSVSVTVH 273

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTUS COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	173.5	11.4	969	2	US-08-284-941-2
2	173.5	11.4	969	2	US-08-447-642-2
3	173.5	11.4	969	3	US-09-236-503-2
4	173.5	11.4	969	5	PCT-US93-02147A-2
5	162	10.7	799	2	US-08-525-940-23
6	162	10.7	799	2	US-08-376-838-23
7	162	10.7	881	2	US-08-525-940-21
8	162	10.7	881	2	US-08-376-838-21
9	162	10.7	915	2	US-08-525-940-18
10	162	10.7	915	2	US-08-376-838-18
11	162	10.7	915	4	US-09-214-555B-2
12	162	10.7	915	4	US-09-214-555B-7
13	161	10.6	288	1	US-08-368-852-15
14	158.5	10.5	379	4	US-09-307-794A-4
15	158.5	10.5	379	4	US-09-305-125A-4
16	158.5	10.5	379	4	US-09-302-775A-4
17	157.5	10.4	288	2	US-08-525-940-15
18	157.5	10.4	288	2	US-08-376-838-15
19	150.5	9.9	380	4	US-09-205-258-441
20	149	9.8	568	1	US-07-862-021B-14
21	149	9.8	568	5	PCT-US93-03164-18
22	148.5	9.8	2523	1	US-08-185-432-14
23	148.5	9.8	2523	4	US-08-899-232-3
24	147.5	9.7	802	1	US-07-862-021B-12
25	147.5	9.7	802	1	US-08-313-288B-12
26	147.5	9.7	802	5	PCT-US93-03164-12
27	146.5	9.7	807	4	US-09-132-769-1

28	146.5	9.7	807	4	US-09-640-173-186	Sequence 186, Appl
29	146.5	9.7	807	4	US-09-713-550-186	Sequence 186, Appl
30	146	9.6	1068	1	US-08-537-210A-2	Sequence 2, Appli
31	146	9.6	1068	3	US-09-113-825-2	Sequence 2, Appli
32	146	9.6	2556	1	US-08-185-432-17	Sequence 17, Appli
33	146	9.6	2556	4	US-08-899-232-2	Sequence 2, Appli
34	144	9.5	807	4	US-09-132-769-3	Sequence 3, Appli
35	143.5	9.5	807	1	US-07-862-021B-10	Sequence 10, Appl
36	143.5	9.5	807	1	US-08-313-288B-10	Sequence 10, Appl
37	143.5	9.5	807	4	US-09-132-769-5	Sequence 5, Appli
38	143.5	9.5	807	5	PCT-US93-03164-10	Sequence 103, Appl
39	139	9.2	366	3	US-08-857-076-103	Sequence 5, Appli
40	139	9.2	486	3	US-08-746-559A-5	Sequence 4, Appli
41	139	9.2	516	3	US-08-746-559A-4	Sequence 2, Appli
42	139	9.2	1367	2	US-08-249-887C-2	Sequence 2, Appli
43	139	9.2	1367	2	US-08-625-819-2	Sequence 2, Appli
44	139	9.2	1367	3	US-08-746-559A-2	Sequence 2, Appli
45	139	9.2	1367	4	US-08-864-641B-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-08-284-941-2
; Sequence 2, Application US/08284941
; Patent No. 5863756
; GENERAL INFORMATION:
; APPLICANT: BARR, PHILIP J
; APPLICANT: KIEFER, MICHAEL C
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND
; TITLE OF INVENTION: PACE 4.1 GENE AND POLYPEPTIDES IN CELLS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
; STREET: FIVE PALO ALTO SQUARE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/284,941
; FILING DATE: 2 August 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: NEELLY PH.D., RICHARD L.
; REGISTRATION NUMBER: 30092
; REFERENCE/DOCKET NUMBER: CHIR-009/01US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 843-5070
; TELEFAX: (415) 857-0663
; TELEX: 380816 COOLY FA
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 969 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-284-941-2

Query Match 11.4%; Score 173.5; DB 2; Length 969;
Best Local Similarity 28.8%; Pred. No. 1.2e-06;
Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;
QY 19 YIGSQNASRRQRMRMHPNVQCGGCATCSD--YNGCLSCXPRLPFALERIGMKQIGV 76
Db 737 YFGDTAARCR-----CHKGCTCSSRAATQCLSCR-RGPF-----HHQSMNT 779

QY 77 CLSSCPGVYGYTRYPDINKTKCKADCTCFNK-NFCTKCKSGFYHLGKCLDNCPEGLE 135
Db 780 CVTLCPAGFYADE--SQKNCLKHPSCKKCVDBPEKCTVCKGFSLARGSCIPDCBPGTY 837
QY 136 ANNHTECVSIVHCEVSENNPWSPTCKGKTG 168
Db 838 FDSLEIRGCECH-----TCG 853

RESULT 2

US-08-447-642-2
; Sequence 2, Application US/08447642
; Patent No. 5989890
; GENERAL INFORMATION:
; APPLICANT: BARR, PHILIP J
; APPLICANT: KIEFER, MICHAEL C
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND
; TITLE OF INVENTION: PACE 4.1 GENE AND POLYPEPTIDES IN CELLS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
; STREET: FIVE PALO ALTO SQUARE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,642
; FILING DATE: 23-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/284,941
; FILING DATE: 2 August 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: NEELEY PH.D., RICHARD L.
; REGISTRATION NUMBER: 30092
; REFERENCE/DOCKET NUMBER: CHIR-009/01US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 843-5070
; TELEFAX: (415) 857-0663
; TELEX: 380816 COOLEY FA
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 969 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-447-642-2

Query Match 11.4%; Score 173.5; DB 2; Length 969;
Best Local Similarity 28.8%; Pred. No. 1.2e-06;
Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;
QY 19 YIGSQNASRRQRMRHPNVSGCGGATCSD--YNGCLSKPRLPFFALERIGMKQIGV 76
Db 737 YFGDTAARRCR-----CHKGCTCSSRAATQCLSCR-RGFY-----HHQEMNT 779
QY 77 CLSSCPGVYGYTRYPDINKTKCKADCTCFNK-NFCTKCKSGFYHLGKCLDNCPEGLE 135
Db 780 CVTLCPAGFYADE--SQKNCLKHPSCKKCVDBPEKCTVCKGFSLARGSCIPDCBPGTY 837
QY 136 ANNHTECVSIVHCEVSENNPWSPTCKGKTG 168
Db 838 FDSLEIRGCECH-----TCG 853

RESULT 3

US-09-236-503-2

; Sequence 2, Application US/09236503
; Patent No. 6277590
; GENERAL INFORMATION:
; APPLICANT: BARR, PHILIP J
; APPLICANT: KIEFER, MICHAEL C
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND 4.1 GENE AND
; TITLE OF INVENTION: Polypeptides in Cells
; FILE REFERENCE: CHIR-009/04US
; CURRENT APPLICATION NUMBER: US/09/236,503
; CURRENT FILING DATE: 1999-01-25
; EARLIER APPLICATION NUMBER: 08/447,642
; EARLIER FILING DATE: 1995-05-23
; EARLIER APPLICATION NUMBER: 08/284,941
; EARLIER FILING DATE: 1994-08-02
; EARLIER APPLICATION NUMBER: 07/848,629
; EARLIER FILING DATE: 1992-03-09
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 969
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-236-503-2
Query Match 11.4%; Score 173.5; DB 3; Length 969;
Best Local Similarity 28.8%; Pred. No. 1.2e-06;
Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;
QY 19 YIGSQNASRRQRMRHPNVSGCGGATCSD--YNGCLSKPRLPFFALERIGMKQIGV 76
Db 737 YFGDTAARRCR-----CHKGCTCSSRAATQCLSCR-RGFY-----HHQEMNT 779
QY 77 CLSSCPGVYGYTRYPDINKTKCKADCTCFNK-NFCTKCKSGFYHLGKCLDNCPEGLE 135
Db 780 CVTLCPAGFYADE--SQKNCLKHPSCKKCVDBPEKCTVCKGFSLARGSCIPDCBPGTY 837
QY 136 ANNHTECVSIVHCEVSENNPWSPTCKGKTG 168
Db 838 FDSLEIRGCECH-----TCG 853

RESULT 4

PCT-US93-02147A-2
; Sequence 2, Application PC/TUS9302147A
; GENERAL INFORMATION:
; APPLICANT: BARR, PHILIP J
; APPLICANT: KIEFER, MICHAEL C
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND
; TITLE OF INVENTION: PACE 4.1 GENE AND POLYPEPTIDES IN CELLS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
; STREET: FIVE PALO ALTO SQUARE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/02147A
; FILING DATE: 19930309
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,629
; FILING DATE: 09-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: NEELEY PH.D., RICHARD L
; REGISTRATION NUMBER: 30092
; REFERENCE/DOCKET NUMBER: CHIR-009/00US

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-7622
; TELEFAX: (415) 857-0663
; TELEX: 380816 COOLEY PA
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 969 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-02147A-2

Query Match 11.4%; Score 173.5; DB 5; Length 969;
Best Local Similarity 28.8%; Pred. No. 1.2e-06;
Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;

QY 19 YIGSQNASRRRRMHNNVSGCGGCGATCSD--YNGCLSKPRLPFFALERIGMKQIGV 76
DB 737 YFDYRARRCR-----CHKGCTCSSRAATQCLSCR-RGFY-----HHQEMNT 779
QY 77 CLSSCPSGYGYTRYPDINKCTKCKADCTCFNK-NFCTKCKSGFYHLGKCLDNCPEGLE 135
DB 780 CVTLCPAGFYADE--SQKNCLKHPSCCKVDRPKCTVCKEGFSLARGSCIPDCRGTY 837
QY 136 ANNHTMECVSIHVCEVSEWNPSPCTKXGKTG 168
DB 838 FDSLEIRGCEHH-----TCG 853

RESULT 5
US-08-525-940-23
; Sequence 23, Application US/08525940
; Patent No. 5866351
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; APPLICANT: Miranda, Luis R.
; TITLE OF INVENTION: CD4+ T-LYMPHOCTE PROTEASES AND GENES
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,940
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/368,852
; FILING DATE: 01-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/088,322
; FILING DATE: 07-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2848-11-C1
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 799 amino acids
; TYPE: amino acid

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-525-940-23

Query Match 10.7%; Score 162; DB 2; Length 799;
Best Local Similarity 22.3%; Pred. No. 9.4e-06;
Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-525-940-23

Query Match 10.7%; Score 162; DB 2; Length 799;
Best Local Similarity 22.3%; Pred. No. 9.4e-06;
Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;

QY 21 GSONASRRRRMHNNVSGCGGCGATCSDYNG--CLSKPRLPFFALERIGMKQIGVCL 78
DB 563 GHYHADK-KRCRCAPN-----CESCFGSHGDCQMSCKYGYFL-----NEETNSCV 607
QY 79 SSCPSG-YGYTRYPDINKCTKCKADCTCFNK-NFCTKCKSGFYHLGKCLDNCPEGLE 137
DB 608 THCPDGSYQDTKK---NLCKKSENCKTCTEPHNCCTCRDGLSLQSGRCVSVCSDGRYFN 664
QY 138 NHTME-----CVS-----IVHCEVSEW-----NPMSPCTKK 163
DB 665 GDCQCFHRPCATCAGAGAGCINCTEGYFMDGRCVQSCSISYYFDHSSSENGYKSKCKC 724
QY 164 GKTG-----GFKRGTE-----TRVREIIQHPSAKGNLCPTNETRKTCTV 202
DB 725 DISCLTCNGGPFKNCTSCPSGYLLDLGMCQGAICKDATBESWABGGFCMLVKKNLNC-- 782
QY 203 QRKCKQK 209
DB 783 QRKVLQK 789

RESULT 6
US-08-976-838-23
; Sequence 23, Application US/08976838
; Patent No. 5981259
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; TITLE OF INVENTION: CD4+ T-LYMPHOCTE PROTEASE NUCLEIC ACID
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,838
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2848-11-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 799 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-976-838-23

Query Match 10.7%; Score 162; DB 2; Length 799;
Best Local Similarity 22.3%; Pred. No. 9.4e-06;
Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;
```


QY 138 NHTME-----CVS-----IVHCEVSEW-----NPWSPTCK 163
Db 747 GDCQCPCHRCATCAGAGAGCINCTEGYPMWDRGCVQSCSISYYFDHSSSENGYKSKKC 806
QY 164 GKTC-----GPKRGTE-----TRVREIIQHPSAKGNLCPPPTNETRKTCTV 202
Db 807 DISCLTCNGPGFKNCTSCPSGYLLDLGMCQGAICKDATESWAEAGGFCMLVKKNLNC-- 864
QY 203 ORKKCQK 209
Db 865 QKVLQO 871

RESULT 9

US-08-525-940-18
; Sequence 18, Application US/08525940
; Patent No. 5866351
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; APPLICANT: Miranda, Luis R.
; APPLICANT: Wolf, Joseph R.
; TITLE OF INVENTION: CD4+ T-LYMPHOCTE PROTEASES AND GENES
; TITLE OF INVENTION: ENCODING SAID PROTEASES
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,940
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/366,852
; FILING DATE: 01-JAN-1995
; PRIOR APPLICATION DATA: US 08/088,322
; APPLICATION NUMBER: US 08/088,322
; FILING DATE: 07-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2848-11-C1
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-525-940-18

Query Match 10.7%; Score 162; DB 2; Length 915;
Best Local Similarity 22.3%; Pred. No. 1.1e-05;
Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;
QY 21 GSONASRRGRORRHPNVSGCGGATCSDYNG--CLSKPRLFPALERIGMKQIGVCL 78
Db 679 GHYHADK-KKCRKCAPN-----CBSCFGSHGDCQMSCKYGYFL-----NEETNSCV 723
QY 79 SSCPSG-YGTRYPDINKTKKADCDTCFNKFCFKSGFYHLGLKCLDNCPEGLEAN 137
Db 724 THCPDGSYQDTKK--NLCKKSENCKTCTBFHNTCTCRDGLSLQSGRCSVSCDGRYFN 780

QY 138 NHTME-----CVS-----IVHCEVSEW-----NPWSPTCK 163
Db 781 GDCQCPCHRCATCAGAGAGCINCTEGYPMWDRGCVQSCSISYYFDHSSSENGYKSKKC 840
QY 164 GKTC-----GPKRGTE-----TRVREIIQHPSAKGNLCPPPTNETRKTCTV 202
Db 841 DISCLTCNGPGFKNCTSCPSGYLLDLGMCQGAICKDATESWAEAGGFCMLVKKNLNC-- 898
QY 203 ORKKCQK 209
Db 899 QKVLQO 905

RESULT 10

US-08-976-838-18
; Sequence 18, Application US/08976838
; Patent No. 5981259
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; TITLE OF INVENTION: CD4+ T-LYMPHOCTE PROTEASE NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,838
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2848-11-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-976-838-18

Query Match 10.7%; Score 162; DB 2; Length 915;
Best Local Similarity 22.3%; Pred. No. 1.1e-05;
Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;
QY 21 GSONASRRGRORRHPNVSGCGGATCSDYNG--CLSKPRLFPALERIGMKQIGVCL 78
Db 679 GHYHADK-KKCRKCAPN-----CBSCFGSHGDCQMSCKYGYFL-----NEETNSCV 723

QY 79 SSCPSG-YGTRYPDINKTKKADCDTCFNKFCFKSGFYHLGLKCLDNCPEGLEAN 137
Db 724 THCPDGSYQDTKK--NLCKKSENCKTCTBFHNTCTCRDGLSLQSGRCSVSCDGRYFN 780
QY 138 NHTME-----CVS-----IVHCEVSEW-----NPWSPTCK 163
Db 781 GDCQCPCHRCATCAGAGAGCINCTEGYPMWDRGCVQSCSISYYFDHSSSENGYKSKKC 840
QY 164 GKTC-----GPKRGTE-----TRVREIIQHPSAKGNLCPPPTNETRKTCTV 202
Db 841 DISCLTCNGPGFKNCTSCPSGYLLDLGMCQGAICKDATESWAEAGGFCMLVKKNLNC-- 898

```
QY 203 QKCKQK 209
Db 899 QKVLQK 905

RESULT 11
US-09-214-555B-2
; Sequence 2, Application US/09214555B
; Patent No. 6380171
; GENERAL INFORMATION:
; APPLICANT: INSTITUT DE RECHERCHE CLINIQUE DE MONTREAL
; TITLE OF INVENTION: PRO-PROTEIN CONVERTING ENZYME
; FILE REFERENCE: PRO-PROTEIN CONVER ENZ
; CURRENT APPLICATION NUMBER: US/09/214,555B
; CURRENT FILING DATE: 1999-01-04
; PRIOR APPLICATION NUMBER: 60/021,008
; PRIOR FILING DATE: 1996-07-26
; PRIOR APPLICATION NUMBER: 2,203,745
; PRIOR FILING DATE: 1997-04-25
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-555B-2

Query Match 10.7%; Score 162; DB 4; Length 915;
Best Local Similarity 22.3%; Pred. No. 1.1e-05;
Matches 55; Conservative 33; Mismatches 81; Indels 76; Gaps 12;

QY 21 GSONASRGRRORRMPNVSQGGCGCATCSDYNG--CLSCPKRLFFALERIGMKQIGVCL 78
Db 679 GHYHADK-KRCKCAPN-----CESCFGSHGDCMSCKYGYFL-----NEETNSCV 723

QY 79 SSCPSG-YGTRYPDINKTKKADCDTCFNKFNCTKCKSGFYHLGKCLDNCBPBGLAN 137
Db 724 THCPDGSQDTRK--NLCKRCSNCKYCTEFHNCTECDGLSLOGSRCSVSCDGRYFN 780

QY 138 NHTMB-----CVS-----IVHCEVSEW-----NPWSPCTKK 163
Db 781 GDCQCPCHRFATCAGAGADGCTEGYFMDGRCVQSCISYTFDHSSENGYKSKKC 840

QY 164 GKTG-----GFKRGTE-----TRVREIIQHPSAKGNLCPTTNETRKCTV 202
Db 841 DISCLTCNGPGPKNCTSCPSGYLLDLGMCQGAICKDATEESWAGGFCMLVKKNLNC-- 898

QY 203 QKCKQK 209
Db 899 QKVLQK 905

RESULT 12
US-09-214-555B-7
; Sequence 7, Application US/09214555B
; Patent No. 6380171
; GENERAL INFORMATION:
; APPLICANT: INSTITUT DE RECHERCHE CLINIQUE DE MONTREAL
; TITLE OF INVENTION: PRO-PROTEIN CONVERTING ENZYME
; FILE REFERENCE: PRO-PROTEIN CONVER ENZ
; CURRENT APPLICATION NUMBER: US/09/214,555B
; CURRENT FILING DATE: 1999-01-04
; PRIOR APPLICATION NUMBER: 60/021,008
; PRIOR FILING DATE: 1996-07-26
; PRIOR APPLICATION NUMBER: 2,203,745
; PRIOR FILING DATE: 1997-04-25
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-555B-7

Query Match 10.6%; Score 161; DB 1; Length 288;
Best Local Similarity 28.2%; Pred. No. 3.6e-06;
Matches 46; Conservative 25; Mismatches 52; Indels 40; Gaps 11;

QY 36 PNVSQ-GCGC-GCATCSDYNGCLSCPKRLFFALERIGMKQIGVCLSSCSGYYGTGTPDI 93
Db 899 QKVLQK 905
```

Db 13 PECSEVCGDGPDPCHND---CL---HYKKL---NNTRICVSSCPGPHY---HADK 58
Qy 94 NKCTKACDCTCPNK--NFTCKCKSGFYH--LGKCLDNCPEGLBANNHTMBCVSIHVC 149
Db 59 KRCKKCAPNCESCFGSHGDCQMSCKYGYFLNEETNSCVTHCPDGSQDTTKNLC-----112
Qy 150 EVSEWNPWSCTKKGKTC-GPKRGTRVREIIHQHPSAKGNLC 191
Db 113 -----RKSENKXKCTBFHXCTECR-----DGLSLQGSRC 142

RESULT 14

US-09-907-794A-4

; Sequence 4, Application US/09907794A
; Patent No. 663468
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Deenoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,794A
; CURRENT FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22 US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095

; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 4
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-907-794A-4

Query Match

10.5%; Score 158.5; DB 4; Length 379;

Best Local Similarity 24.9%; Pred. No. 8.1e-06;

Matches 60; Conservative 72; Mismatches 72; Indels 83; Gaps 16;

Qy 42 CQGGCAT---CSDYNGC-----LSCKPRLFFALERIGMKQIGVCLSS---C 81
Db 182 CPGCRNGGFCNERRICCEPDGPHGPHCEKALCTPRCHG-----GLCVTFPGCIC 232
Qy 82 PSGYGYTRYPDINKCTRC-KADCD-TCFNKNFCTKCKSGFYHLGKCLDNCPEGLBANNH 139
Db 233 PPGFYG-----VNCCKANCSTTCFNGTC-----FY--PKCI--CPPGLEGB-- 271
Qy 140 TMECVSIVHCEVSEWNPWSCTKKGKTCGFKRGTRVREIIHQHPSAKGNLCPP----- 193
Db 272 -----QCEISKCP--QPCRNGKCGIG---KSKCKSGYQGLDCKSKVCEPCCGAHG 318
Qy 194 -TNETKCTVQRKKCKGKRGKGRERKRK-----KPNKGSKEAIPDSLSLESSKEIP 246
Db 319 TCHEPNKC-----QCQGWGHRHCKRYEASLIHALRPAQLRQHTPSLKAERRDPP 373
Qy 247 E 247
Db 374 E 374

RESULT 15

US-09-905-125A-4

; Sequence 4, Application US/09905125A

; Patent No. 6664376

; GENERAL INFORMATION:

; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Deenoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905,125A

```
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 4
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-905-125A-4

Query Match      10.5%; Score 158.5; DB 4; Length 379;
Best Local Similarity 24.9%; Pred. No. 8.1e-06;
Matches 60; Conservative 26; Mismatches 72; Indels 83; Gaps 16;

Qy 42 CQGCAT---CSDYNGC-----LSCKPLFPALERIGMKQIGVCLSS-----C 81
Db 182 CPGGCRNGGFCNERICECPDGFHGHPCBKALCTPCNNG-----GLCVTPGFCIC 232
Qy 82 PSGYTGTRYPDINKTKC-KAPCD-TCFNNEFTCKKSGFYHLGKCLDNCPEGLEANNH 139
Db 233 PGFYTG-----VNCDKANCSTTCFNGGTC-----FY--PGKCI--CPPGLEGE-- 271
Qy 140 TMECVSIHVCVSEWNPSPCTKKKTCGKGTETRVREIIQHPSAKGNLCPP----- 193
Db 272 -----QCEISKCP--QPCNNGKCIQ-----KSKCKSGYGGDLCSKPKVCEPCGGAHG 318
Qy 194 -TNETRKCTVQRKKQKGERKKERKK-----KPNKGSKEAIPDSKSLESKEIP 246
Db 319 TCHEPNKC-----QCEGWHGRHCNRYEASLIHALRPAGALRQHTPSLKAERDRDP 373
Qy 247 E 247
Db 374 E 374
```

Search completed: June 29, 2004, 17:04:01
Job time : 16.2464 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 17:02:26 ; Search time 36.7755 Seconds
(Without alignments)
2098.641 Million cell updates/sec

Title: US-09-894-912A-13

Perfect score: 1516

Sequence: 1 MGHRLISWLFILNFWYIGSNASGRGRQRMRHFNVSQCGGCATCSDYNGCLSKCP 273

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1166195 seqs, 282705291 residues

Total number of hits satisfying chosen parameters: 1166195

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2_5/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_5/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1516	100.0	273	10	US-09-894-912A-13
2	1505	99.3	272	10	US-09-894-912A-10
3	1505	99.3	272	10	US-09-894-912A-34
4	1505	99.3	272	14	US-10-125-852-23
5	1472	97.1	265	10	US-09-894-912A-26
6	1472	97.1	292	9	US-09-745-763-166
7	1472	97.1	292	10	US-09-894-912A-48
8	1472	97.1	292	14	US-10-125-852-25
9	1400	92.3	251	10	US-09-894-912A-16
10	1310.5	86.4	279	10	US-09-894-912A-32
11	1275	84.1	225	14	US-10-185-770-4
12	1213	80.0	239	12	US-10-087-192-1782
13	1178	77.7	239	15	US-10-094-886-172
14	1111	73.3	195	15	US-10-094-886-176
15	1107	73.0	195	15	US-10-094-886-174

16	981	64.7	180	12	US-10-087-192-1779	Sequence 1779, Ap
17	903	59.6	160	10	US-09-894-912A-14	Sequence 14, Appl
18	903	59.6	160	12	US-10-276-174-1744	Sequence 1744, Ap
19	656	43.3	263	14	US-10-125-852-18	Sequence 18, Appl
20	648.5	42.8	243	14	US-10-125-852-21	Sequence 21, Appl
21	644	42.5	265	14	US-10-125-852-24	Sequence 24, Appl
22	638	42.1	229	10	US-09-894-912A-25	Sequence 25, Appl
23	626.5	41.3	236	15	US-10-108-260A-4829	Sequence 4829, Ap
24	584.5	38.6	243	14	US-10-185-770-2	Sequence 2, Appl
25	577.5	38.1	243	14	US-10-125-852-13	Sequence 13, Appl
26	577.5	38.1	243	16	US-10-467-042-12	Sequence 12, Appl
27	573	37.8	250	14	US-10-125-852-3	Sequence 3, Appl
28	545	35.9	222	14	US-10-125-852-15	Sequence 15, Appl
29	540.5	35.7	229	14	US-10-125-852-6	Sequence 6, Appl
30	400.5	26.4	190	12	US-10-087-192-873	Sequence 873, App
31	365.5	24.1	161	12	US-10-087-192-876	Sequence 876, App
32	276	18.2	46	10	US-09-894-912A-18	Sequence 18, Appl
33	234.5	15.5	131	14	US-10-125-852-9	Sequence 9, Appl
34	221	14.6	42	10	US-09-894-912A-22	Sequence 22, Appl
35	213	14.1	37	10	US-09-894-912A-20	Sequence 20, Appl
36	202	13.3	110	14	US-10-125-852-11	Sequence 11, Appl
37	186.5	12.3	1548	14	US-10-180-903-2	Sequence 2, Appl
38	177.5	11.7	43	14	US-10-125-852-7	Sequence 7, Appl
39	173.5	11.4	969	10	US-09-961-403-6	Sequence 6, Appl
40	160.5	10.6	337	12	US-10-357-820-20	Sequence 20, Appl
41	160.5	10.6	343	12	US-10-357-820-8	Sequence 8, Appl
42	160.5	10.6	365	12	US-10-357-820-4	Sequence 4, Appl
43	160.5	10.6	365	12	US-10-357-820-6	Sequence 6, Appl
44	160.5	10.6	373	12	US-10-357-820-10	Sequence 10, Appl
45	160.5	10.6	479	9	US-09-764-898-221	Sequence 221, App

ALIGNMENTS

RESULT 1

US-09-894-912A-13
; Sequence 13, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; FACTOR-LIKE POLYPEPTIDES AND POLYNUCLEOTIDES
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; CURRENT FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-894-912A-13

Query Match 100.0%; Score 1516; DB 10; Length 273;
Best Local Similarity 100.0%; Pred. No. 1.6e-115;
Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MGHRLISWLFILNFWYIGSNASGRGRQRMRHFNVSQCGGCATCSDYNGCLSKCP	60
Db	1	MGHRLISWLFILNFWYIGSNASGRGRQRMRHFNVSQCGGCATCSDYNGCLSKCP	60
Qy	61	RLPFALERIGMKIGVCLSSCPSGYGYTRYPTDINKCTCKKADCTCFNKNPCTCKSGFY	120

Db 61 RLFLALERIGKQIGVCLSSCPGSGYGYTRYPDINKTKCADCDTCFNKNFCTCKSGFY 120
Qy 121 LHLGKLDNCPGLEANNHTMECVSIHVCEVSEWNPSPCTKGGKTCGFKGTETRVREI 180
Db 121 LHLGKLDNCPGLEANNHTMECVSIHVCEVSEWNPSPCTKGGKTCGFKGTETRVREI 180
Qy 181 IOHPSAKGNLCPTNETRKTCTVQRKCKQGERGKGRKRRKKNPKNGESKEAIPDSKSL 240
Db 181 IOHPSAKGNLCPTNETRKTCTVQRKCKQGERGKGRKRRKKNPKNGESKEAIPDSKSL 240
Qy 241 SSKPEORENKQOKKRRKVDKQKSVSVTVH 273
Db 241 SSKPEORENKQOKKRRKVDKQKSVSVTVH 273

RESULT 2

US-09-894-912A-10
; Sequence 10, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; PRIOR FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-894-912A-10

Query Match 99.3%; Score 1505; DB 10; Length 272;
Best Local Similarity 100.0%; Pred. No. 1.3e-114;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 HRLISWLFILNFMFMEYIGSQNASRGRQRHHPNVSQCGGCATCSDYNGCLCKPRL 62
Db 2 HRLISWLFILNFMFMEYIGSQNASRGRQRHHPNVSQCGGCATCSDYNGCLCKPRL 61
Qy 63 PFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKTKCADCDTCFNKNFCTCKSGFY 122
Db 62 PFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKTKCADCDTCFNKNFCTCKSGFY 121
Qy 123 LGKLDNCPGLEANNHTMECVSIHVCEVSEWNPSPCTKGGKTCGFKGTETRVREI 182
Db 122 LGKLDNCPGLEANNHTMECVSIHVCEVSEWNPSPCTKGGKTCGFKGTETRVREI 181
Qy 183 HPSAKGNLCPTNETRKTCTVQRKCKQGERGKGRKRRKKNPKNGESKEAIPDSKSL 242
Db 182 HPSAKGNLCPTNETRKTCTVQRKCKQGERGKGRKRRKKNPKNGESKEAIPDSKSL 241
Qy 243 KEIPEORENKQOKKRRKVDKQKSVSVTVH 273
Db 242 KEIPEORENKQOKKRRKVDKQKSVSVTVH 272

RESULT 3

US-09-894-912A-34
; Sequence 34, Application US/09894912A

; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; PRIOR FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-894-912A-34

Query Match 99.3%; Score 1505; DB 10; Length 272;
Best Local Similarity 100.0%; Pred. No. 1.3e-114;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 HRLISWLFILNFMFMEYIGSQNASRGRQRHHPNVSQCGGCATCSDYNGCLCKPRL 62
Db 2 HRLISWLFILNFMFMEYIGSQNASRGRQRHHPNVSQCGGCATCSDYNGCLCKPRL 61
Qy 63 PFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKTKCADCDTCFNKNFCTCKSGFY 122
Db 62 PFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKTKCADCDTCFNKNFCTCKSGFY 121
Qy 123 LGKLDNCPGLEANNHTMECVSIHVCEVSEWNPSPCTKGGKTCGFKGTETRVREI 182
Db 122 LGKLDNCPGLEANNHTMECVSIHVCEVSEWNPSPCTKGGKTCGFKGTETRVREI 181
Qy 183 HPSAKGNLCPTNETRKTCTVQRKCKQGERGKGRKRRKKNPKNGESKEAIPDSKSL 242
Db 182 HPSAKGNLCPTNETRKTCTVQRKCKQGERGKGRKRRKKNPKNGESKEAIPDSKSL 241
Qy 243 KEIPEORENKQOKKRRKVDKQKSVSVTVH 273
Db 242 KEIPEORENKQOKKRRKVDKQKSVSVTVH 272

RESULT 4

US-10-125-852-23
; Sequence 23, Application US/10125852
; Publication No. US20030032034A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH FACTOR-1
; FILE REFERENCE: HYS-43A
; CURRENT APPLICATION NUMBER: US/10/125,852
; PRIOR FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: US 60/316,368
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: US 09/799,451
; PRIOR FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-125-852-23

Query Match 99.3%; Score 1505; DB 14; Length 272;
Best Local Similarity 100.0%; Pred. No. 1.3e-114;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HURLISWLFILNFMFYIGSNASRRRRRHPNVSQCGGCATCSYNGCLCKPRL 62
DB 2 HURLISWLFILNFMFYIGSNASRRRRRHPNVSQCGGCATCSYNGCLCKPRL 61

QY 63 PFALERIGHKQIGVCLSSCPGSGYGYTRYPDINKTKCKADCDTCFKNKPFCTKCKSGFYHL 122
DB 62 PFALERIGHKQIGVCLSSCPGSGYGYTRYPDINKTKCKADCDTCFKNKPFCTKCKSGFYHL 121

QY 123 LGKCLDNCPEGLNHNTHMECVSIHVCYSEWNPSPCTKKGTCGPKGTETRVREIIQ 182
DB 122 LGKCLDNCPEGLNHNTHMECVSIHVCYSEWNPSPCTKKGTCGPKGTETRVREIIQ 181

QY 183 HPSAKGNLCPPPTNETRKCTVQRKKCKGGRKKRKKRKKPKNKGSKKAI PDSKSLSS 242
DB 182 HPSAKGNLCPPPTNETRKCTVQRKKCKGGRKKRKKRKKPKNKGSKKAI PDSKSLSS 241

QY 243 KEIPEORENKKQKKRKVDKQKSVSVTVH 273
DB 242 KEIPEORENKKQKKRKVDKQKSVSVTVH 272

RESULT 5
US-09-894-912A-26
; Sequence 26, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; FILE OF INVENTION: FACTOR-LIKE POLYPEPTIDES AND POLYNUCLEOTIDES
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; CURRENT FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 26
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-894-912A-26

Query Match 97.1%; Score 1472; DB 10; Length 265;
Best Local Similarity 100.0%; Pred. No. 5.9e-112;
Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HURLISWLFILNFMFYIGSNASRRRRRHPNVSQCGGCATCSYNGCLCKPRL 62
DB 2 HURLISWLFILNFMFYIGSNASRRRRRHPNVSQCGGCATCSYNGCLCKPRL 61

QY 63 PFALERIGHKQIGVCLSSCPGSGYGYTRYPDINKTKCKADCDTCFKNKPFCTKCKSGFYHL 122
DB 62 PFALERIGHKQIGVCLSSCPGSGYGYTRYPDINKTKCKADCDTCFKNKPFCTKCKSGFYHL 121

QY 123 LGKCLDNCPEGLNHNTHMECVSIHVCYSEWNPSPCTKKGTCGPKGTETRVREIIQ 182
DB 122 LGKCLDNCPEGLNHNTHMECVSIHVCYSEWNPSPCTKKGTCGPKGTETRVREIIQ 181

QY 183 HPSAKGNLCPPPTNETRKCTVQRKKCKGGRKKRKKRKKPKNKGSKKAI PDSKSLSS 242
DB 182 HPSAKGNLCPPPTNETRKCTVQRKKCKGGRKKRKKRKKPKNKGSKKAI PDSKSLSS 241

Db 182 HPSAKGNLCPPPTNETRKCTVQRKKCKGGRKKRKKRKKPKNKGSKKAI PDSKSLSS 241

QY 243 KEIPEORENKKQKKRKVDKQKSVSVTVH 272
DB 242 KEIPEORENKKQKKRKVDKQKSVSVTVH 272

RESULT 6
US-09-745-763-166
; Sequence 166, Application US/09745763
; Patent No. US20020065394A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; McCoy, John M.
; Lavallie, Edward R.
; Collins-Racie, Lisa A.
; Evans, Cheryl
; Merberg, David
; Treacy, Maurice
; Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 219
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/745,763
; FILING DATE: 18-Jun-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 166:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 292 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 166:
US-09-745-763-166

Query Match 97.1%; Score 1472; DB 9; Length 292;
Best Local Similarity 100.0%; Pred. No. 6.6e-112;
Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HURLISWLFILNFMFYIGSNASRRRRRHPNVSQCGGCATCSYNGCLCKPRL 62
DB 2 HURLISWLFILNFMFYIGSNASRRRRRHPNVSQCGGCATCSYNGCLCKPRL 61

QY 63 PFALERIGHKQIGVCLSSCPGSGYGYTRYPDINKTKCKADCDTCFKNKPFCTKCKSGFYHL 122
DB 62 PFALERIGHKQIGVCLSSCPGSGYGYTRYPDINKTKCKADCDTCFKNKPFCTKCKSGFYHL 121

QY 123 LGKCLDNCPEGLNHNTHMECVSIHVCYSEWNPSPCTKKGTCGPKGTETRVREIIQ 182
DB 122 LGKCLDNCPEGLNHNTHMECVSIHVCYSEWNPSPCTKKGTCGPKGTETRVREIIQ 181

QY 183 HPSAKGNLCPPPTNETRKCTVQRKKCKGGRKKRKKRKKPKNKGSKKAI PDSKSLSS 242
DB 182 HPSAKGNLCPPPTNETRKCTVQRKKCKGGRKKRKKRKKPKNKGSKKAI PDSKSLSS 241

Db 182 HPSAKGNLCPPNTRKCTVQRKKCKQGBRGKKRRRRKKRKKPNKGSKEAIPDSKSLESS 241
Qy 243 KEIPEORENKQKKRKKVQDKQS 266
Db 242 KEIPEORENKQKKRKKVQDKQS 265

RESULT 7
US-09-894-912A-48
; Sequence 48, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; FILE OF INVENTION: FACTOR-LIKE POLYPEPTIDES AND POLYNUCLEOTIDES
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; CURRENT FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 48
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-894-912A-48

Query Match 97.1%; Score 1472; DB 10; Length 292;
Best Local Similarity 100.0%; Pred. No. 6.6e-112;
Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 HURLISWLFILNFMFMEYIGSQNASRRRRRRHPNVSQCGGCATCSDYNGCLCKPRL 62
Db 2 HURLISWLFILNFMFMEYIGSQNASRRRRRRHPNVSQCGGCATCSDYNGCLCKPRL 61
Qy 63 PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKCTCKKADCTCFNKNFCTCKSGFYH 122
Db 62 PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKCTCKKADCTCFNKNFCTCKSGFYH 121
Qy 123 LKCLDNCPEGLEANNHTECVSIHVCEVSEWNPSPCTKKGKTCGFKGTETRVREIIQ 182
Db 122 LKCLDNCPEGLEANNHTECVSIHVCEVSEWNPSPCTKKGKTCGFKGTETRVREIIQ 181
Qy 183 HPSAKGNLCPPNTRKCTVQRKKCKQGBRGKKRRRRKKRKKPNKGSKEAIPDSKSLESS 242
Db 182 HPSAKGNLCPPNTRKCTVQRKKCKQGBRGKKRRRRKKRKKPNKGSKEAIPDSKSLESS 241
Qy 243 KEIPEORENKQKKRKKVQDKQS 266
Db 242 KEIPEORENKQKKRKKVQDKQS 265

RESULT 8
US-10-125-852-25
; Sequence 25, Application US/10125852
; Publication No. US20030032034A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH FACTOR-L
; FILE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES
; FILE REFERENCE: HYS-43A
; CURRENT APPLICATION NUMBER: US/10/125,852
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: US 60/316,368

; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: US 09/799,451
; PRIOR FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-125-852-25

Query Match 97.1%; Score 1472; DB 14; Length 292;
Best Local Similarity 100.0%; Pred. No. 6.6e-112;
Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 HURLISWLFILNFMFMEYIGSQNASRRRRRRHPNVSQCGGCATCSDYNGCLCKPRL 62
Db 2 HURLISWLFILNFMFMEYIGSQNASRRRRRRHPNVSQCGGCATCSDYNGCLCKPRL 61
Qy 63 PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKCTCKKADCTCFNKNFCTCKSGFYH 122
Db 62 PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKCTCKKADCTCFNKNFCTCKSGFYH 121
Qy 123 LKCLDNCPEGLEANNHTECVSIHVCEVSEWNPSPCTKKGKTCGFKGTETRVREIIQ 182
Db 122 LKCLDNCPEGLEANNHTECVSIHVCEVSEWNPSPCTKKGKTCGFKGTETRVREIIQ 181
Qy 183 HPSAKGNLCPPNTRKCTVQRKKCKQGBRGKKRRRRKKRKKPNKGSKEAIPDSKSLESS 242
Db 182 HPSAKGNLCPPNTRKCTVQRKKCKQGBRGKKRRRRKKRKKPNKGSKEAIPDSKSLESS 241
Qy 243 KEIPEORENKQKKRKKVQDKQS 266
Db 242 KEIPEORENKQKKRKKVQDKQS 265

RESULT 9
US-09-894-912A-15
; Sequence 15, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; FILE OF INVENTION: FACTOR-LIKE POLYPEPTIDES AND POLYNUCLEOTIDES
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; CURRENT FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-894-912A-16

Query Match 92.3%; Score 1400; DB 10; Length 251;
Best Local Similarity 100.0%; Pred. No. 4.1e-106;
Matches 251; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 23 QNASRRRRRRHPNVSQCGGCATCSDYNGCLCKPRLFPALERIGMKQIGVCLSSCP 82
Db 1 QNASRRRRRRHPNVSQCGGCATCSDYNGCLCKPRLFPALERIGMKQIGVCLSSCP 60

Qy	83	SGYVGYTRYPDINKCTCKKADCDTCFNKNPCTCKCKSGPYHLGKCLDNCPEGLEANNHTME	142
Db	61	SGYVGYTRYPDINKCTCKKADCDTCFNKNPCTCKCKSGPYHLGKCLDNCPEGLEANNHTME	120
Qy	143	CVSIVHCEVSWNPWSPCTCKGKTCGPKRGTYETRVREIIOHPSAKGNI.CPPTNETRKCTV	202
Db	121	CVSIVHCEVSWNPWSPCTCKGKTCGPKRGTYETRVREIIOHPSAKGNI.CPPTNETRKCTV	180
Qy	203	QRKXCKQGERGCKGRERKRKXPNKGBSKEAIPDSKSLSSSKEIPBQRENKQOQKCKRVQD	262
Db	181	QRKXCKQGERGCKGRERKRKXPNKGBSKEAIPDSKSLSSSKEIPBQRENKQOQKCKRVQD	240
Qy	263	KQKSVSVSVSTVH	273
Db	241	KQKSVSVSVSTVH	251

```

RESULT 10
US-09-894-912A-32
; Sequence 32, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; TITLE OF INVENTION: FACTOR-LIKE POLYPEPTIDES AND POLYNUCLEOTIDES
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; CURRENT FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 32
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-894-912A-32

```

RESULT 11
US-10-185-770-4

```

; Sequence 4, Application US/10185770
; Publication No. US20030022217A1
; GENERAL INFORMATION:
; APPLICANT: CECCARDI, Toni et al.
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL0001247
; CURRENT APPLICATION NUMBER: US/10/185,770
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/301,852
; PRIOR FILING DATE: 2001-07-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-185-770-4

Query Match      84.1%; Score 1275; DB 14; Length 225;
Best Local Similarity 100.0%; Pred. No. 5.5e-96;
Matches 224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      3  HRLRLSWLFILNFMVEYIGSONASRGRRQRMHPNVVSQGGCGGCATCSDYNGCLSCPKRL 62
Db      2  HRLRLSWLFILNFMVEYIGSONASRGRRQRMHPNVVSQGGCGGCATCSDYNGCLSCPKRL 61

Qy      63  PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKCTCKACDCTCFNKNFCTCKCKSGFYLH 122
Db      62  PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKCTCKACDCTCFNKNFCTCKCKSGFYLH 121

Qy      123  LGKCLDNCPLGLEANNHMTMCVSIHVCEVSEMPNWPSPCTKKGKTCGFRGTETVRRIIQ 182
Db      122  LGKCLDNCPLGLEANNHMTMCVSIHVCEVSEMPNWPSPCTKKGKTCGFRGTETVRRIIQ 181

Qy      183  HPSAKGNLCPTNTRKCTVORKKCOXGRRGKGRERKQKPNK 226
Db      182  HPSAKGNLCPTNTRKCTVORKKCOXGRRGKGRERKQKPNK 225

```

```

RESULT 12
US-10-087-192-1782
; Sequence 1782, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: fastSEQ for Windows Version 4.0
; SEQ ID NO 1782
; LENGTH: 239
; TYPE: FRT
; ORGANISM: Homo sapiens
US-10-087-192-1782

```

Query Match	80.0%	Score 1213	DB 12	Length 239
Best Local Similarity	97.3%	Pred. No. 6.6e-91		
Matches 214	Conservative 1	Mismatches 5	Indels 0	Gaps 0

Qy	3	HLRLISWLFILINPMWEYIGSONASRGRRRRRHPNVISQCGGCGATCSDYNGCLSKPRL	62
nb	19	HLPLISWLFILINPMWEYIGSONASRGRRRRRHPNVISQCGGCGATCSDYNGCLSKPRL	78

QY 63 PFALERIGMKQIGVCLSSCPGSGYGYGTRYPDINKCTKACDCTCFWKNFCTKCKSGFYLA 122
Db 79 PFALERIGMKQIGVCLSSCPGSGYGYGTRYPDINKCTKACDCTCFWKNFCTKCKSGFYLA 138
QY 123 LKGLCLDNCPGLEANNHTMECVSIHVCEVSEWNPSPCTKKGKTCGFKRGTTETRVREIIQ 182
Db 139 LKGLCLDNCPGLEANNHTMECVSIHVCEVSEWNPSPCTKKGKTCGFKRGTTETRVREIIQ 198
QY 183 HPSAKGNLCPPPTNETRKCTVQRKKCKGGRKKGRERK 222
Db 199 HPSAKGNLCPPPTNETRKCTVQRKKCKGGRGTIIIGEEKK 238

RESULT 13

US-10-094-886-172
; Sequence 172, Application US/10094886
; Publication No. US20040002120A1
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Patturajan, Meera
; APPLICANT: Burgess, Catherine
; APPLICANT: Vernet, Corine A.
; APPLICANT: Li, Li
; APPLICANT: Gorman, Linda
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Boldog, Ferenc
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shenoy, Suresh
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Taupier, Raymond J., Jr.
; APPLICANT: Miller, Charles
; APPLICANT: Casman, Stacie
; APPLICANT: Pena, Carol
; APPLICANT: Gangolli, Bsha
; APPLICANT: Gusev, Vladimir
; APPLICANT: Smithson, Glennda
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Gerlach, Valerie
; APPLICANT: Pochart, Pascal
; APPLICANT: Fernandes, Elma
; APPLICANT: Shinkets, Richard
; APPLICANT: Rastelli, Luca
; APPLICANT: Spaderma, Steven
; APPLICANT: Larochelle, William
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD

; FILE REFERENCE: 21402-290 B
; CURRENT APPLICATION NUMBER: US/10/094,886
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/274,322
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/313,182
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/288,052
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/318,510
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/314,018
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/296,693
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/313,626
; PRIOR FILING DATE: 2001-08-21
; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 298
; SOFTWARE: Patentin 2.1
; SEQ ID NO 172
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-886-172

Query Match 77.7%; Score 1178; DB 15; Length 239;
Best Local Similarity 97.7%; Pred. No. 4.6e-88;
Matches 210; Conservative 0; Mismatches 1; Indels 4; Gaps 2;

QY 3 HURLISWFLIILNFMEYIGSONASRRGRORRMRHPNVSQCGCGCATCSDYNGCLSCKPRL 62
Db 2 HURLISWFLIILNFMEYIGSONASRRGRORRMRHPNVSQCGCGCATCSDYNGCLSCKPRL 61
QY 63 PFALERIGMKQIGVCLSSCPGSGYGYGTRYPDINKCT-KC---KADCTCFWKNFCTKCKSG 118
Db 62 PFALERIGMKQIGVCLSSCPGSGYGYGTRYPDINKCTSKPHEKADCTCFWKNFCTKCKSG 121
QY 119 FYHLGKCLDNCPGLEANNHTMECVSIHVCEVSEWNPSPCTKKGKTCGFKRGTTETRV 178
Db 122 FYHLGKCLDNCPGLEANNHTMECVSVHCEVSEWNPSPCTKKGKTCGFKRGTTETRV 181
QY 179 EIIQHPSAKGNLCPPPTNETRKCTVQRKKCKGGRG 213
Db 182 EIIQHPSAKGNLCPPPTNETRKCTVQRKKCKGGRG 216

RESULT 14

US-10-094-886-176
; Sequence 176, Application US/10094886
; Publication No. US20040002120A1
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Patturajan, Meera
; APPLICANT: Burgess, Catherine
; APPLICANT: Vernet, Corine A.
; APPLICANT: Li, Li
; APPLICANT: Gorman, Linda
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Boldog, Ferenc
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shenoy, Suresh
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Taupier, Raymond J., Jr.
; APPLICANT: Miller, Charles
; APPLICANT: Casman, Stacie
; APPLICANT: Pena, Carol
; APPLICANT: Gangolli, Bsha
; APPLICANT: Gusev, Vladimir
; APPLICANT: Smithson, Glennda
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Gerlach, Valerie
; APPLICANT: Pochart, Pascal
; APPLICANT: Fernandes, Elma
; APPLICANT: Shinkets, Richard
; APPLICANT: Rastelli, Luca
; APPLICANT: Spaderma, Steven
; APPLICANT: Larochelle, William
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-290 B
; CURRENT APPLICATION NUMBER: US/10/094,886
; CURRENT FILING DATE: 2002-03-07
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/274,322
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/313,182
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/288,052

;; PRIOR FILING DATE: 2001-05-02
;; PRIOR APPLICATION NUMBER: 60/318,510
;; PRIOR FILING DATE: 2001-09-10
;; PRIOR APPLICATION NUMBER: 60/274,281
;; PRIOR FILING DATE: 2001-03-08
;; PRIOR APPLICATION NUMBER: 60/314,018
;; PRIOR FILING DATE: 2001-08-21
;; PRIOR APPLICATION NUMBER: 60/274,194
;; PRIOR FILING DATE: 2001-03-08
;; PRIOR APPLICATION NUMBER: 60/274,849
;; PRIOR FILING DATE: 2001-03-09
;; PRIOR APPLICATION NUMBER: 60/296,693
;; PRIOR FILING DATE: 2001-06-07
;; PRIOR APPLICATION NUMBER: 60/313,626
;; PRIOR FILING DATE: 2001-08-21
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 298
;; SOFTWARE: PatentIn 2.1
;; SEQ ID NO 176
;; LENGTH: 195
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-094-886-176

Query Match 73.3%; Score 1111; DB 15; Length 195;
Best Local Similarity 100.0%; Pred. No. 1.1e-82;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 GSNASRGRRQRHHPNVSGCGGCATCSDYNGCLSCPKRLFPALERIGMKQIGVCLSS 80
DB 1 GSNASRGRRQRHHPNVSGCGGCATCSDYNGCLSCPKRLFPALERIGMKQIGVCLSS 60

QY 81 CPSGGYGTYPDINKCTCKKADCDTCFNKNFCTKCKSGFYHLGKCLDNCPEGLEANNHT 140
DB 61 CPSGGYGTYPDINKCTCKKADCDTCFNKNFCTKCKSGFYHLGKCLDNCPEGLEANNHT 120

QY 141 MCVSVIVHCEVSEWNPWSPCTKKGKTCGPKRGTTETRVREIIQHPSAKGNLCPTNETRKC 200
DB 121 MCVSVIVHCEVSEWNPWSPCTKKGKTCGPKRGTTETRVREIIQHPSAKGNLCPTNETRKC 180

QY 201 TVQRKKCKQKRG 213
DB 181 TVQRKKCKQKRG 193

RESULT 15
US-10-094-886-174
;; Sequence 174, Application US/10094886
;; Publication No. US20040002120A1
;; GENERAL INFORMATION:
;; APPLICANT: Kekuda, Ramesh
;; APPLICANT: Tchernev, Velizar T.
;; APPLICANT: Liu, Xiaohong
;; APPLICANT: Spytek, Kimberly A.
;; APPLICANT: Patturajan, Meera
;; APPLICANT: Burgess, Catherine
;; APPLICANT: Vernet, Corine A.
;; APPLICANT: Li, Li
;; APPLICANT: Gorman, Linda
;; APPLICANT: Malyankar, Uriel M.
;; APPLICANT: Boldog, Ferenc
;; APPLICANT: Guo, Xiaojia
;; APPLICANT: Shenoy, Suresh
;; APPLICANT: Padigar, Muralidhara
;; APPLICANT: Taupier, Raymond J., Jr.
;; APPLICANT: Miller, Charles
;; APPLICANT: Casman, Stacie
;; APPLICANT: Pena, Carol
;; APPLICANT: Gangolli, Esha
;; APPLICANT: Gusev, Vladimir
;; APPLICANT: Smithson, Glennda
;; APPLICANT: Zerhusen, Bryan
;; APPLICANT: Gerlach, Valerie

;; APPLICANT: Pochart, Pascal
;; APPLICANT: Fernandes, Eima
;; APPLICANT: Shinkets, Richard
;; APPLICANT: Rastelli, Luca
;; APPLICANT: Spaderna, Steven
;; APPLICANT: LaRochele, William
;; APPLICANT: Zhong, Wei
;; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD;
;; FILE REFERENCE: 21402-290 B
;; CURRENT APPLICATION NUMBER: US/10/094,886
;; CURRENT FILING DATE: 2002-03-07
;; PRIOR APPLICATION NUMBER: 60/274,322
;; PRIOR FILING DATE: 2001-03-08
;; PRIOR APPLICATION NUMBER: 60/313,182
;; PRIOR FILING DATE: 2001-08-17
;; PRIOR APPLICATION NUMBER: 60/288,052
;; PRIOR FILING DATE: 2001-05-02
;; PRIOR APPLICATION NUMBER: 60/318,510
;; PRIOR FILING DATE: 2001-09-10
;; PRIOR APPLICATION NUMBER: 60/274,281
;; PRIOR FILING DATE: 2001-03-08
;; PRIOR APPLICATION NUMBER: 60/314,018
;; PRIOR FILING DATE: 2001-08-21
;; PRIOR APPLICATION NUMBER: 60/274,194
;; PRIOR FILING DATE: 2001-03-08
;; PRIOR APPLICATION NUMBER: 60/274,849
;; PRIOR FILING DATE: 2001-03-09
;; PRIOR APPLICATION NUMBER: 60/296,693
;; PRIOR FILING DATE: 2001-06-07
;; PRIOR APPLICATION NUMBER: 60/313,626
;; PRIOR FILING DATE: 2001-08-21
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 298
;; SOFTWARE: PatentIn 2.1
;; SEQ ID NO 174
;; LENGTH: 195
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-094-886-174

Query Match 73.0%; Score 1107; DB 15; Length 195;
Best Local Similarity 99.5%; Pred. No. 2.2e-82;
Matches 192; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 21 GSNASRGRRQRHHPNVSGCGGCATCSDYNGCLSCPKRLFPALERIGMKQIGVCLSS 80
DB 1 GSNASRGRRQRHHPNVSGCGGCATCSDYNGCLSCPKRLFPALERIGMKQIGVCLSS 60

QY 81 CPSGGYGTYPDINKCTCKKADCDTCFNKNFCTKCKSGFYHLGKCLDNCPEGLEANNHT 140
DB 61 CPSGGYGTYPDINKCTCKKADCDTCFNKNFCTKCKSGFYHLGKCLDNCPEGLEANNHT 120

QY 141 MCVSVIVHCEVSEWNPWSPCTKKGKTCGPKRGTTETRVREIIQHPSAKGNLCPTNETRKC 200
DB 121 MCVSVIVHCEVSEWNPWSPCTKKGKTCGPKRGTTETRVREIIQHPSAKGNLCPTNETRKC 180

QY 201 TVQRKKCKQKRG 213
DB 181 TVQRKKCKQKRG 193

Search completed: June 29, 2004, 17:15:57
Job time : 40.7755 sec

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 16:56:05 ; Search time 13.915 Seconds
(without alignments)
1887.186 Million cell updates/sec

Title: US-09-894-912A-13

Perfect score: 1516

Sequence: 1 MGLRLISWLFILNFMVYI.....QQKRRVQDKQSVSVTVH 273

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: Pir1:*

2: Pir2:*

3: Pir3:*

4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	192	12.7	1299	2	furin (EC 3.4.21.7)
2	186.5	12.3	1548	2	serine proteinase
3	184	12.1	962	2	subtilisin-like pr
4	184	12.1	975	2	subtilisin-like pr
5	176.5	11.6	1680	2	furin (EC 3.4.21.7)
6	173.5	11.4	969	1	subtilisin-like pr
7	168	11.1	915	2	probable propotei
8	167.5	11.0	932	2	PAC34A - mouse (fr
9	167	11.0	915	1	subtilisin-like pr
10	162	10.7	899	2	subtilisin-like pr
11	162	10.7	915	2	subtilisin-like pr
12	161	10.6	440	2	hypothetical prote
13	158.5	10.5	379	2	Gene PAC34 protein
14	156.5	10.3	937	2	P-spondin precursor
15	156	10.3	803	2	protein bli-4D [im
16	153.5	10.1	942	2	Notch B protein -
17	150.5	9.9	1203	2	insulin-like growt
18	149	9.8	1371	2	Xotch protein - Af
19	148.5	9.8	2524	2	cell-fate determin
20	147.5	9.7	2471	2	probable kexin (EC
21	145	9.6	570	2	hypothetical prote
22	143.5	9.6	1620	2	P-spondin - rat
23	143.5	9.5	807	2	Wnt inhibitory fac
24	142	9.4	378	2	insulin-like growt
25	142	9.4	677	2	C42125
26	139	9.2	1367	1	IGHRU1
27	139	9.2	2555	2	notch protein homo
28	138.5	9.1	1369	2	protein-tyrosine k
29	138	9.1	2531	2	notch protein homo

30 138 9.1 2531 2 A46019 notch-1 protein -
31 137 9.0 2437 2 S42612 transmembrane prot
32 136 9.0 540 2 B47417 insulin receptor-r
33 135.5 8.9 1382 1 INHUR insulin receptor p
34 134.5 8.9 1111 2 T26972 hypothetical prote
35 134.5 8.9 1372 2 A34157 insulin receptor p
36 134.5 8.9 1383 2 A36080 insulin receptor p
37 129.5 8.5 2101 2 S57245 insulin receptor (
38 129.5 8.5 2148 1 A56081 insulin receptor -
39 128 8.4 327 2 A46484 apoptosis-mediati
40 128 8.4 1268 2 B36502 insulin receptor-r
41 127 8.4 861 2 A48825 Notch homolog Motc
42 126.5 8.3 3871 2 T22812 hypothetical prote
43 125.5 8.3 667 2 A48579 trophozoite surfac
44 125 8.2 837 2 S43656 furin (EC 3.4.21.7
45 124 8.2 1274 2 T42017 cysteine rich prot

ALIGNMENTS

RESULT 1

T43251

furin (EC 3.4.21.75) - fall armyworm

N;Alternate names: paired basic amino acid cleaving enzyme; proprotein convertase; serin
C;Species: Spodoptera frugiperda (fall armyworm)

C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000

C;Accession: T43251

R;Cleplik, M.; Klenk, H.

submitted to the EMBL Data Library, January 1996

A;Description: Cloning and functional characterization of FURIN from Spodoptera frugiper

A;Reference number: Z22368

A;Accession: T43251

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1299 <CIB>

A;Cross-references: EMBL:Z68888; NID:g1167859; PID:e219690; PIDN:CAA93116.1

A;Experimental source: Clone Sfurin 6; ovary

C;Function:

A;Description: responsible for the endoproteolytic processing of proproteins with specifi

C;Keywords: hydrolase; serine proteinase

Query Match 12.7%; Score 192; DB 2; Length 1299;

Best Local Similarity 27.8%; Pred. No. 2.6e-05;

Matches 63; Conservative 26; Mismatches 78; Indels 60; Gaps 12;

QY 37 NVSQGCGGCGATCD--NKNFTCKSGGFFLHGKLDNCPGELHANNHTMVCVSIYHCEVSE 153

Db 795 SVCRPCAHCATCSEADGCTSCHEHL-----VLHDGTCWASCPSPHYET---BDDM 843

QY 96 CTYCKKADCTCF--NKNFTCKSGGFFLHGKLDNCPGELHANNHTMVCVSIYHCEVSE 153

Db 844 CARKHSCDTQCGPGTQCVTCHPSTVALDGRCVTSCPPAYADKKRKC---MRCVPVG- 899

QY 154 WNPWSPTCKGKTGCGFRTGTETVRETIQHPSAKGNLCPP---TNETKRC-TVQKKCKOK 209

Db 900 ---CSTCT-----SAPFLSCRPAWELNKKKGKCMVSGDKCSA 933

QY 210 GERGKGRKRKKPKNK-----GESKE---AIPDSKSLSSKEIPE 247

Db 934 GBFAV---DQKCRKCPACDSCYGENEGHCLTCFNPPLLQDYKCVPE 977

RESULT 2

S34583

serine proteinase (EC 3.4.21.-) PC6B - mouse

C;Species: Mus musculus (house mouse)

C;Date: 02-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 05-Nov-1999

C;Accession: S34583

R;Nakagawa, T.; Murakami, K.; Nakayama, K.

FEBS Lett. 327, 165-171, 1993

A;Title: Identification of an isoform with an extremely large Cys-rich region of PC6, a

A;Reference number: S34583; MUID:93327934; PMID:8335106

A:Accession: S34583
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1548 <NAK>
A:Cross-references: GB:D17583; NID:G407344; PIDN:BAA04507.1; PID:dl005033; PID:G440374
C:Keywords: hydrolase; serine proteinase

Query Match 12.1%; Score 186.5; DB 2; Length 1548;
Best Local Similarity 25.6%; Pred. No. 6.8e-05;
Matches 66; Conservative 31; Mismatches 90; Indels 77; Gaps 15;

QY 39 SQCGGCGATCSDYNG--CLSKPRFLPFFALRIGMKQIGVCLSSCPGSGYGYTRYPI--N 94
DB 880 NQCHSSCKTC---NSGLASCPTGYLWLO-----ACVPSGCG-----TWPSVTSG 924
QY 95 KCTKCKACDCTCNKPNCTKCKS----GFYLHLGKCLDNCPEGLRANNTMB-CVSIIVHC 149
DB 925 SCEKSEDCVSCGADLCQQLSQPDNTLLHGRCTVHSCPEGFYAKDGVCEHCSS--PC 982
QY 150 EVSEWNPSPCTKKG-----KTC-----GPKRGITRTVREIILQHPGSAKGNL- 190
DB 983 KTCGNATSCNSCEGDFVLHGVCMKTCPEKHVAVGVCKHCPERCQDCIHEKTKCKCMP 1042
QY 191 -----CPPT--NETRKTCTVQRKKCKQKGERGKGRKRKKPKNKGSKKAIPDS 236
DB 1043 DFPLNDMCHRSCKPXFYPMRQCVPCHKNC-----LRCNGPKEDDCKVCADTS 1091
QY 237 KSLSS---KEIPE-----QRENKQ 253
DB 1092 KALLHNGCLDCECPGTYKBEENDS 1115

RESULT 3

JCS571
subtilisin-like proprotein convertase (EC 3.4.21.-) PACE4 precursor, splice form E-II -
C:Species: Homo sapiens (man)
C:Date: 23-Sep-1997 #sequence_revision 23-Sep-1997 #text_change 20-Jun-2000
C:Accession: JCS571
R:Morii, K.; Kii, S.; Teuji, A.; Nagahama, M.; Imanaki, A.; Hayashi, K.; Akamatsu, T.; Na
J. Biochem. 121, 941-948, 1997
A:Title: A novel human PACE4 isoform, PACE4B is an active processing protease containing
A:Reference number: JCS570; MUID:97335942; PMID:9192737
A:Accession: JCS571
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-962 <MOR>
A:Cross-references: DDBJ:D87994; NID:G2330550; PIDN:BAA21792.1; PID:G2330551
A:Experimental source: brain cerebellum
C:Comment: This enzyme is a processing protease and responsible for processing of various
ch it is retained intracellularly.

C:Genetics:
A:Gene: GDB:PACE4
A:Cross-references: GDB:131390; OMIM:167405
A:Map position: 15q26-15q26
A:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
C:Keywords: glycoprotein; hydrolase; serine proteinase
F:1-62/Domain: signal sequence #status predicted <SIG>
F:63-149/Domain: propeptide #status predicted <PRO>
F:196-434/Domain: subtilisin homology <SBT>
F:938-954/Domain: hydrophobic cluster #status predicted <HCL>
F:205,246,347,420/Active site: Asp, His, Asn, Ser #status predicted
F:259/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 12.1%; Score 184; DB 2; Length 962;
Best Local Similarity 26.6%; Pred. No. 7e-05;
Matches 62; Conservative 24; Mismatches 87; Indels 60; Gaps 14;
QY 19 YIGSQNASRGRQRMRHNPVSGCGGCGATCSD--YNGCLSKPRFLPFFALERIGMKQIGV 76
DB 724 YFGDTAARCR-----CHKGCTCSSRAATQCLSCR-RGFI-----HHQEMNT 766
QY 77 CLSSCPGSGYGYTRYPDINKTKKADCTCFNK-NFCTKCKGPFYHLGKCLDNCPEGLE 135

Query Match 12.1%; Score 184; DB 2; Length 962;
Best Local Similarity 26.6%; Pred. No. 7e-05;
Matches 62; Conservative 24; Mismatches 87; Indels 60; Gaps 14;

QY 19 YIGSQNASRGRQRMRHNPVSGCGGCGATCSD--YNGCLSKPRFLPFFALERIGMKQIGV 76
DB 724 YFGDTAARCR-----CHKGCTCSSRAATQCLSCR-RGFI-----HHQEMNT 766
QY 77 CLSSCPGSGYGYTRYPDINKTKKADCTCFNK-NFCTKCKGPFYHLGKCLDNCPEGLE 135

DB 767 CVTLCPAGFYADB--SQKNCLKCHPCKKCVDFPEKCTVCKBGFSLARGSCIPDCBGTY 824
QY 136 ANNHTMEC-----VSIVHCEVS--EMNPSPCTKKGTCGPKRGTTETRVRE 179
DB 825 FDSLIROGECCHTCTGTCVGPGRBECIHCANPHFHDW-----KCVPAAGBGFYPE----- 875
QY 180 IIOHPSAKGNLC-----PPTNETRKTCTVQRKKCKQKGERGKGRKRKKPKNKG 228
DB 876 --EMPGLPHKVCRRYPPGGE-RQATVS-----SKGVPG--GQSLSSASSPAGE 919

RESULT 4

JCS570
subtilisin-like proprotein convertase (EC 3.4.21.-) PACE4 precursor, splice form E-I -
C:Species: Homo sapiens (man)
C:Date: 23-Sep-1997 #sequence_revision 23-Sep-1997 #text_change 20-Jun-2000
C:Accession: JCS570
R:Morii, K.; Kii, S.; Teuji, A.; Nagahama, M.; Imanaki, A.; Hayashi, K.; Akamatsu, T.; Na
J. Biochem. 121, 941-948, 1997
A:Title: A novel human PACE4 isoform, PACE4B is an active processing protease containing
A:Reference number: JCS570; MUID:97335942; PMID:9192737
A:Accession: JCS570
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-975 <MOR>
A:Cross-references: DDBJ:D87993; NID:G2330548; PIDN:BAA21791.1; PID:G2330549
A:Experimental source: brain cerebellum
C:Comment: This enzyme is a processing protease and responsible for processing of various
ch it is retained intracellularly.

C:Genetics:
A:Gene: GDB:PACE4
A:Cross-references: GDB:131390; OMIM:167405
A:Map position: 15q26-15q26
A:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
C:Keywords: alternative splicing; glycoprotein; hydrolase; serine proteinase
F:1-62/Domain: signal sequence #status predicted <SIG>
F:63-149/Domain: propeptide #status predicted <PRO>
F:196-434/Domain: subtilisin homology <SBT>
F:952-968/Domain: hydrophobic cluster #status predicted <HCL>
F:205,246,347,420/Active site: Asp, His, Asn, Ser #status predicted
F:259/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 12.1%; Score 184; DB 2; Length 975;
Best Local Similarity 26.6%; Pred. No. 7.1e-05;
Matches 62; Conservative 24; Mismatches 87; Indels 60; Gaps 14;
QY 19 YIGSQNASRGRQRMRHNPVSGCGGCGATCSD--YNGCLSKPRFLPFFALERIGMKQIGV 76
DB 737 YFGDTAARCR-----CHKGCTCSSRAATQCLSCR-RGFI-----HHQEMNT 779
QY 77 CLSSCPGSGYGYTRYPDINKTKKADCTCFNK-NFCTKCKGPFYHLGKCLDNCPEGLE 135
DB 780 CVTLCPAGFYADB--SQKNCLKCHPCKKCVDFPEKCTVCKBGFSLARGSCIPDCBGTY 837
QY 136 ANNHTMEC-----VSIVHCEVS--EMNPSPCTKKGTCGPKRGTTETRVRE 179
DB 838 FDSLIROGECCHTCTGTCVGPGRBECIHCANPHFHDW-----KCVPAAGBGFYPE----- 888
QY 180 IIOHPSAKGNLC-----PPTNETRKTCTVQRKKCKQKGERGKGRKRKKPKNKG 228
DB 889 --EMPGLPHKVCRRYPPGGE-RQATVS-----SKGVPG--GQSLSSASSPAGE 932

RESULT 5

A43434
furin (EC 3.4.21.75) 2 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 31-Mar-2000
C:Accession: A43434
R:Roebroek, A.J.; Creemers, J.W.; Pauli, I.G.; Kurzik-Dumke, U.; Rentrop, M.; Gateff, B.
J. Biol. Chem. 267, 17208-17215, 1992
A:Title: Cloning and functional expression of Dfurin2, a subtilisin-like proprotein proc
A:Reference number: A43434; MUID:92381036; PMID:1512259

A:Accession: A43434
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1680 <ROE>
A:Cross-references: GB:M94375; NID:g157461; PID:g157462
A:Note: sequence extracted from NCBI backbone (NCBIN:111933, NCBIP:111934)
C:Genetics:
A:Gene: FlyBase:Pur2
A:Cross-references: FlyBase:FBgn0004598
C:Superfamily: subtilisin homology
C:Keywords: hydrolase; serine proteinase; transmembrane protein
F:409-652/Domain: subtilisin homology <SBT>
F:418,457,638/Active site: Asp, His, Ser #status predicted

Query Match 11.6%; Score 176.5; DB 2; Length 1680;
Best Local Similarity 28.0%; Pred. No. 0.00033;
Matches 60; Conservative 24; Mismatches 77; Indels 53; Gaps 12;

QY 3 HLRISWLFILNPMVEYIGSQASRGRRMRHNPVSGQGCGATCSDY-NGCLSCKPR 61
DB 1038 HLHVID-LAVCLQPCPGYFENS-----RNTCVF-----CEPNCASQDHPYCTSCDHH 1087
QY 62 LFFALERIGMKQIGVCLSSCPSTGYTRYPDINKCTKADCDTCF--NRNFTCKKSGF 119
DB 1088 LVMHEHK-----CYSACPLDYET--EDNKCACFCHSTCATCNGPTDQDCITCSSR 1136
QY 120 YHLGKCLDNCPEGLEANNHMECVSIHVCEVSEWNPSPCTKKGKTCGFRGTETRV-R 178
DB 1137 YAWQNKILSCPDPGFYADKRLCM-----PCQEGCKTC-----TSNGVCS 1177
QY 179 EIIQHPGAKGNLCPTTNETRKTQVKK-CQKGE 211
DB 1178 ECLQNT-----LNRDKCIVSGEGCGESE 1203

RESULT 6
A39490
subtilisin-like proprotein convertase (EC 3.4.21.-) PACE4 precursor, splice form A - human
N:Alternate names: kexin homology
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 31-Mar-2000
C:Accession: A39490
R:Kiefer, M.C.; Tucker, J.E.; Joh, R.; Landeberg, K.E.; Saltman, D.; Barr, P.J.
DNA Cell Biol. 10, 757-769, 1991
A:Title: Identification of a second human subtilisin-like protease gene in the fes/fps region
A:Reference number: A39490; MUID:92075167; PMID:1741956
A:Molecule type: mRNA
A:Residues: 1-969 <KIE>
A:Cross-references: GB:M80482; NID:g189531; PIDN:AAAS9998.1; PID:g189532
C:Genetics:
A:Gene: GDB:PACE4
A:Cross-references: GDB:131390; OMIM:167405
A:Map position: 15q26-15q26
C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
C:Keywords: alternative splicing; hydrolase; serine proteinase
F:150-969/Product: serine proteinase PACE4 #status predicted <STG>
F:196-434/Domain: subtilisin homology <SBT>
F:205,246,420/Active site: Asp, His, Ser #status predicted

Query Match 11.4%; Score 173.5; DB 1; Length 969;
Best Local Similarity 28.8%; Pred. No. 0.00034;
Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;

QY 19 YIGSQASRGRRMRHNPVSGQGCGATCSD--YNGCLSCKPRLPFALERIGMKQIGV 76
DB 737 YFGDTAARRCR-----CHKGETCSSRAATQCLSCR-RGFY-----HHQEMNT 779
QY 77 CLSCSPSTGYTRYPDINKCTKADCDTCFNK-NFTCKKSGFYHLGKCLDNCPEGLE 135
DB 780 CVTLCPAGFYADB--SQNKCLKHPCKCKVDEPEKCTVCKEGFSLARGSCIPDCPGTY 837
QY 136 ANNHITMECVSIHVCEVSEWNPSPCTKKGKTCG 168

DB 838 FDSLEIRGCGCHH-----TCG 853

RESULT 7
B48225
probable proprotein convertase (EC 3.4.21.-) 5 precursor - rat
N:Alternate names: PCS precursor
C:Species: Rattus norvegicus (Norway rat)
C:Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 23-Feb-1997
C:Accession: B48225
R:Lusson, J.; Vieau, D.; Hamelin, J.; Day, R.; Chretien, M.; Seidah, N.G.
Proc. Natl. Acad. Sci. U.S.A. 90, 6691-6695, 1993
A:Title: cDNA structure of the mouse and rat subtilisin/kexin-like PCS: a candidate proprotein
A:Reference number: A48225; MUID:93342056; PMID:8341687
A:Accession: B48225
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-915 <LUS>
A:Cross-references: GB:L14933
C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
C:Keywords: duplication; glycoprotein; hydrolase; integrin binding; serine proteinase
F:1-34/Domain: signal sequence #status predicted <SIG>
F:35-915/Product: probable proprotein convertase 5 #status predicted <PRO>
F:117-915/Domain: probable proprotein convertase 5 #status experimental <MAT>
F:164-402/Domain: subtilisin homology <SBT>
F:173,214,388/Active site: Asp, His, Ser #status predicted

Query Match 11.1%; Score 168; DB 2; Length 915;
Best Local Similarity 26.3%; Pred. No. 0.00075;
Matches 49; Conservative 25; Mismatches 54; Indels 58; Gaps 11;

QY 36 PNVSQ-GCOG-----GCATCDYNGCLSCKPRLPFALERIGMKQIGVCLSSCPSTGYTRY 90
DB 640 PRCSEVCGDGPDPDHTCDLHYHKLNNTR-----ICVSSCPGHPF---H 682
QY 91 PDINKCTKADCDTCFNKFN--CTKCKSGFYLR--LGKCLDNCPEGLEANNHMECVSI 146
DB 683 ADKRCCKCAPNCSCFSGSHADQCLSKYGYFLNEETSSCVAQCEGSGYQIKKNIC--- 739
QY 147 VHCVEWNPSPCTKKGKTC-GFRGTETRVREIIQHPGAKGNLCPTTNETRKTQVOK 205
DB 740 -----GKSCENKCTCTGFHNCTE-----CKGGL---SLQGSRCVS--- 771
QY 206 KQKGE 211
DB 772 TCSDGQ 777

RESULT 8
I52527
PACE4A - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 16-Jul-1999
C:Accession: I52527
R:Hosaka, M.; Murakami, K.; Nakayama, K.
Biomed. Res. 15, 383-390, 1994
A:Title: PACE4A is a ubiquitous endoprotease that has similar but not identical substrates
A:Reference number: I52527
A:Accession: I52527
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-932 <RES>
A:Cross-references: GB:D50060; NID:g769700; PIDN:BA08777.1; PID:g769701
C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
F:172-410/Domain: subtilisin homology <SBT>

Query Match 11.0%; Score 167.5; DB 2; Length 932;
Best Local Similarity 24.1%; Pred. No. 0.00082;
Matches 52; Conservative 25; Mismatches 78; Indels 61; Gaps 11;

QY 19 YIGSQASRGRRMRHNPVSGQGCGATCSDYN--GCLSCKPRLPFALERIGMKQIGV 76

Db 700 YFGDAARRCR-----CHKGCTCTGRSPAQLCSR-RGFY-----HHQBTWT 742

 Qy 77 CLSCPSGGYYTRYPDINKCKKADCDTPFNKFNFKTGSGFYHLGLKCLDNCPEGLE 135
 :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 743 CVTLCPAGLYADESRL--CLRCHPSCOKCVDPERKCTVCKEGFSLARGSCIPIDEPGTY 800
 :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|

 Qy 136 ANNHTEC-----VSIVHCEVS---EWNPWSPCTKKGTGCVPKRGFTET 176
 :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 801 FDSELVKGCGCHHTCRTGVGSRECIHCASKSFHFQDMKCVPACGE----GF----- 848
 :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|

 Qy 177 VRIIHOHSASGNLCPPNTNETRCTVQRKKOOGKER 212
 :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 849 -----YPEMPGL--PHKVRCBENCLSCGSSR 876
 :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|

RESULT 9
A48225
subtilisin-like proprotein convertase (EC 3.4.21.-) PC5 precursor - mouse
N;Alternate names: kexin homolog; serine proteinase PC6
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: A48225; JX0248
R;Lusson, J.; Vieau, D.; Hamelin, J.; Day, R.; Chretien, M.; Seidah, N.G.
Proc. Natl. Acad. Sci. U.S.A. 90, 6691-6695, 1993
A;Title: cDNA structure of the mouse and rat subtilisin/kexin-like PC5: a candidate proteolytic enzyme from the brain of the house mouse.
A;Reference number: A48225; PMID:93342056; PMID:8341687
A;Accession: A48225
A>Status: preliminary
A:Molecule type: mRNA
A;Residues: 1-915 <JUS>
A;Cross-references: GB:L14932; NID:G293327; PIDN:AAA74636.1; PID:G293328
R:Nakagawa, T.; Hosaka, M.; Torii, S.; Watanabe, T.; Murakami, K.; Nakayama, K.
J. Biochem. 113, 132-135, 1993
A;Title: Identification and functional expression of a new member of the mammalian kex2 gene family.
A;Reference number: JX0248; PMID:93224489; PMID:8468318
A;Accession: JX0248
A:Molecule type: mRNA
A;Residues: 1-915 <NAK>
A;Cross-references: DBM:D12619; NID:g220565; PIDN:BAA02143.1; PID:g220566
A>Note: The authors translated the codon GGC for residue 915 as Ala
C;Superfamily: subtilisin-like proteinase PACE; subtilisin homology
C;Keywords: duplication; glycoprotein; hydrolase; integrin binding; serine proteinase
P;1-34/Domain: signal sequence #status predicted <SIG>
P;35-116/Domain: propetide #status predicted <PRO>
P;117-915/Product: proprotein convertase PC5 #status experimental <MAT>
P;164-402/Domain: subtilisin homology <SET>
P;173,214,388/Active site: Asp, His, Ser #status predicted

Query Match 11.0%; Score 167; DB 1; Length 915;
Best Local Similarity 23.4%; Pred. No. 0.00089;
Matches 58; Conservative 29; Mismatches 81; Indels 80; Gaps 13;

Qy 21 GSQNASKRRORMRHPNVSOCCGGCATCSDNYG--CLSKPLFFALEIGHMKQTGVL 78
 :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 679 GHYHADK-KRKRCAPN-----CESCFGSHGDQCISCKGYFL-----NEETSVCV 723
 :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|

 Qy 79 SSCPSGGYYTRYPDINK--CTKCADCDTCNNPFCTKSGFYHLGLKCLDNCPEGLE 136
 :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 724 TCPCDG-----SYEDIKNVGCKSENCAICIFHNCTCKGSLQSRSVTVCBEDGPFP 779
 :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|

 Qy 137 NNHTMB-----CVS-----IVHCEVSEW-----NPWSPTCK 162
 :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 780 NGHDQPCHRFCATCSGAGDCINCTEGYWBEGRVCQSCSVSYLLDSHGEGYSKR 839
 :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|

 Qy 163 KGKTG-----GPKR-----GT---ETRVRERIIOHPSAKNLCPPTNETRKTCT 201
 :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 840 CDNSCLTNGPGFRNKCSSCPGYLLDLGTQMGAICKDATEESWAEGGFVKLNKLNC- 898
 :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|

 Qy 202 VORKKOQ 209
 :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 899 --QRKVLQQ 905
 :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|

F-spondin precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 17-Nov-2000
C:Accession: A47723
R:Ruiz i Altaba, A.; Cox, C.; Jessell, T.M.; Klar, A.
Proc. Natl. Acad. Sci. U.S.A. 90, 8268-8272, 1993
A:Title: Ectopic neural expression of a floor plate marker in frog embryos injected with A:Reference number: A47723; MUID:93376785; PMID:8367492
A:Accession: A47723
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-803 <RUI>
A:Cross-references: GB:L09123; NID:G409244; PID:AAA19105.1; PID:G409245
C:Superfamily: F-spondin; thrombospondin type 1 repeat homology
F:435-489/Domain: thrombospondin type 1 repeat homology <THR2>
F:607-662/Domain: thrombospondin type 1 repeat homology <THR1>

Query Match	10.3%;	Score 156;	DB 2;	Length 803;
Best Local Similarity	24.6%;	Pred. No. 0.0042;		
Matches	58;	Conservative	35;	Mismatches 91; Indels 52; Gaps 12;

Qy 38 VSQCGCGCATCSYNGCLSCCKPRLPFALERIGHQKIGVCLSSCPS--GYGTGYPDINK 95
Db 544 VNEECPSSCIVTNAEWECS-----ATCMGKKRHRMIKMTPADGSMCKADTTEVEK 598
Qy 96 C--TKCK-----ADCTCFNKNPCTKCKSGFYHLHGKCLDNCPLGEANN--- 138
Db 599 QNMPECHTIPCVLSPWSEWSDCSVTCGKGRTTRQ---MLKSPSELGDCNEELKQVEK 655
Qy 139 -HTWECVSLVHCEVSEWNWPSCTKKGKTCGKRGKTETRVREIIQHPSAKGNLCPTNET 197
Db 656 QMLPECP--ISCELTWYWSSEC--NKSCG--KGHMIRTEMTWEPQGANVPETVQR 708
Qy 198 RKCTVQRKCKQGERGKGRKRKKPNKGESKEAIPDSKLSBSKGIPEQRENKQ 253
Db 709 KKRLL--RKCK-----SSGVRRHLK-----DAREKRSEKIKESDGEQ 747

Search completed: June 29, 2004, 17:03:05
Job time : 15.915 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 16:54:24 ; Search time 8.28277 Seconds
(Without alignments)
1716.232 Million cell updates/sec

Title: US-09-894-912A-13

Perfect score: 1516

Sequence: 1 MGLRLISWLFILNMEVI.....QQKKRKVDKQKSVSVTH 273

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	186.5	12.3	1877	1	PCK5 MOUSE
2	184.5	12.2	1696	1	PCK5 BRACL
3	176.5	11.6	1679	1	FUR2 DROME
4	173.5	11.4	569	1	PAC4 HUMAN
5	168	11.1	1877	1	PCK5 RAT
6	162	10.7	913	1	PCK5 HUMAN
7	158.5	10.5	379	1	WIFI HUMAN
8	157.5	10.4	374	1	WIFI XENLA
9	156.5	10.3	937	1	PAC4 RAT
10	156	10.3	803	1	FSPQ XENLA
11	150.5	9.9	2470	1	NTC2 MOUSE
12	149	9.8	1370	1	IGIR RAT
13	148.5	9.8	2524	1	NOTC XENLA
14	147.5	9.7	2471	1	NTC2 RAT
15	145	9.6	943	1	BL14 CAEEL
16	143.5	9.5	379	1	WIFI MOUSE
17	143.5	9.5	807	1	FSPQ RAT
18	142	9.4	378	1	WIFI BRARE
19	142	9.4	1373	1	IGIR MOUSE
20	139	9.2	1367	1	IGIR HUMAN
21	139	9.2	2556	1	NTC1 HUMAN
22	138	9.1	2531	1	NTC1 RAT
23	138	9.1	2531	1	NTC1 RAT
24	137	9.0	2437	1	NTC1 BRARE
25	135.5	8.9	1382	1	INSR HUMAN
26	134.5	8.9	1372	1	INSR MOUSE
27	134.5	8.9	1383	1	INSR RAT
28	133.5	8.8	2482	1	WVF FIG
29	130.5	8.6	1300	1	IGIR MOUSE
30	130	8.6	2471	1	NTC2 HUMAN
31	129.5	8.5	2146	1	INSR DROME
32	128	8.4	327	1	TNR6 MOUSE
33	128	8.4	1297	1	IGIR HUMAN

RESULT 1
PCK5_MOUSE
ID PCK5_MOUSE STANDARD; PRT; 1877 AA.
AC Q04552; Q62040; 687 1 VS41 GIALA
DT 01-FEB-1995 (Rel. 31, Created) 8.3 687 1 TS11 GIALA
DT 16-OCT-2001 (Rel. 40, Last sequence update) 8.1 461 1 TR1A FIG
DT 15-MAR-2004 (Rel. 43, Last annotation update) 8.1 559 1 TRAP PLAPA
DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-) 8.1 581 1 IRR RAT
DE (Proprotein convertase PCS) (Subtilisin/kexin-like protease PC5) (PC6) 7.9 1429 1 L112 CAEEL
DE (Subtilisin-like proprotein convertase 6) (SPC6). 7.8 1300 1 IRR CAVPO
GN PCSK5. 7.8 1300 1 IRR CAVPO
OS Mus musculus (Mouse). 7.8 2703 1 NOTC DROME
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; 7.8 1477 1 HTK7_HYDAT
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 7.8 867 1 SSPO BOVIN
OX NCBI_TaxID=10090; 7.8 833 1 DL DROME
RN [1] 7.8 833 1 DL DROME
RP SEQUENCE OF 330-1877 FROM N.A. (ISOFORM PC5B). 7.8 1367 1 LT23 CAEEL
RC STRAIN=ICR; TISSUE=Intestine;
RX MEDLINE=93327934; PubMed=8335106;
RA Nakagawa T., Murakami K., Nakayama K.;
RT "Identification of an isoform with an extremely large Cys-rich region of PC6, a Kex2-like processing endoprotease.";
RL FEBS Lett. 327:165-171(1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM PC5A).
RC TISSUE=Brain, and Intestine;
RX MEDLINE=93224489; PubMed=8468318;
RA Nakagawa T., Hosaka M., Torii S., Watanabe T., Murakami K., Nakayama K.;
RT "Identification and functional expression of a new member of the mammalian Kex2-like processing endoprotease family: its striking structural similarity to PACB4.";
RL J. Biochem. 113:132-135(1993).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM PC5A).
RC TISSUE=Adrenal cortex;
RX MEDLINE=93342056; PubMed=8341687;
RA Luessen J., Vieau D., Hamelin J., Day R., Chretien M., Seidah N.G.;
RT "cDNA structure of the mouse and rat subtilisin/kexin-like PC5: a candidate proprotein convertase expressed in endocrine and nonendocrine cells.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).
RN [4]
RP PARTIAL SEQUENCE, AND SUBCELLULAR LOCATION.
RX MEDLINE=97103178; PubMed=8947550;
RA De Bie I., Marcinkiewicz M., Malide D., Lazure C., Nakayama K., Bendayan M., Seidah N.G.;
RT "The isoforms of proprotein convertase PC5 are sorted to different subcellular compartments.";
RL J. Cell Biol. 135:1261-1275(1996).
RN [5]
RP DEVELOPMENTAL EXPRESSION.
RX MEDLINE=96293359; PubMed=898813;
RA Constam D.B., Calton M., Robertson E.J.;
RT "SPC4, SPC6, and the novel protease SPC7 are coexpressed with bone morphogenetic proteins at distinct sites during embryogenesis.";

34 126 8.3 687 1 VS41 GIALA
35 125.5 8.3 687 1 TS11 GIALA
36 123 8.1 461 1 TR1A FIG
37 123 8.1 559 1 TRAP PLAPA
38 123 8.1 581 1 IRR RAT
39 119.5 7.9 1429 1 L112 CAEEL
40 119 7.8 1300 1 IRR CAVPO
41 119 7.8 2703 1 NOTC DROME
42 118.5 7.8 1477 1 HTK7_HYDAT
43 118 7.8 867 1 SSPO BOVIN
44 117.5 7.8 833 1 DL DROME
45 117.5 7.8 1367 1 LT23 CAEEL

ALIGNMENTS

PCK5_MOUSE
ID PCK5_MOUSE STANDARD; PRT; 1877 AA.
AC Q04552; Q62040; 687 1 VS41 GIALA
DT 01-FEB-1995 (Rel. 31, Created) 8.3 687 1 TS11 GIALA
DT 16-OCT-2001 (Rel. 40, Last sequence update) 8.1 461 1 TR1A FIG
DT 15-MAR-2004 (Rel. 43, Last annotation update) 8.1 559 1 TRAP PLAPA
DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-) 8.1 581 1 IRR RAT
DE (Proprotein convertase PCS) (Subtilisin/kexin-like protease PC5) (PC6) 7.9 1429 1 L112 CAEEL
DE (Subtilisin-like proprotein convertase 6) (SPC6). 7.8 1300 1 IRR CAVPO
GN PCSK5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE OF 330-1877 FROM N.A. (ISOFORM PC5B).
RC STRAIN=ICR; TISSUE=Intestine;
RX MEDLINE=93327934; PubMed=8335106;
RA Nakagawa T., Murakami K., Nakayama K.;
RT "Identification of an isoform with an extremely large Cys-rich region of PC6, a Kex2-like processing endoprotease.";
RL FEBS Lett. 327:165-171(1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM PC5A).
RC TISSUE=Brain, and Intestine;
RX MEDLINE=93224489; PubMed=8468318;
RA Nakagawa T., Hosaka M., Torii S., Watanabe T., Murakami K., Nakayama K.;
RT "Identification and functional expression of a new member of the mammalian Kex2-like processing endoprotease family: its striking structural similarity to PACB4.";
RL J. Biochem. 113:132-135(1993).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM PC5A).
RC TISSUE=Adrenal cortex;
RX MEDLINE=93342056; PubMed=8341687;
RA Luessen J., Vieau D., Hamelin J., Day R., Chretien M., Seidah N.G.;
RT "cDNA structure of the mouse and rat subtilisin/kexin-like PC5: a candidate proprotein convertase expressed in endocrine and nonendocrine cells.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).
RN [4]
RP PARTIAL SEQUENCE, AND SUBCELLULAR LOCATION.
RX MEDLINE=97103178; PubMed=8947550;
RA De Bie I., Marcinkiewicz M., Malide D., Lazure C., Nakayama K., Bendayan M., Seidah N.G.;
RT "The isoforms of proprotein convertase PC5 are sorted to different subcellular compartments.";
RL J. Cell Biol. 135:1261-1275(1996).
RN [5]
RP DEVELOPMENTAL EXPRESSION.
RX MEDLINE=96293359; PubMed=898813;
RA Constam D.B., Calton M., Robertson E.J.;
RT "SPC4, SPC6, and the novel protease SPC7 are coexpressed with bone morphogenetic proteins at distinct sites during embryogenesis.";

RL J. Cell Biol. 134:181-191(1996).
 RP [6]
 RP DEVELOPMENTAL EXPRESSION.
 RA MEDLINE=97436919; PubMed=9291583;
 RA Rancourt S.L., Rancourt D.B.;
 RT "Murine subtilisin-like proteinase SP6 is expressed during embryonic
 RT implantation, somitogenesis, and skeletal formation.";
 RL Dev. Genet. 21:75-81(1997).
 CC -!- FUNCTION: Likely to represent a widespread endoprotease activity
 CC within the constitutive and regulated secretory pathway. Capable
 CC of cleavage at the RX(K/R)R consensus motif. May be responsible
 CC for the maturation of gastrointestinal peptides. May be involved
 CC in the cellular proliferation of adrenal cortex via the activation
 CC of growth factors.
 CC -!- CATALYTIC ACTIVITY: Release of mature proteins from their
 CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa
 CC can be any amino acid and Yaa is Arg or Lys.
 CC -!- SUBCELLULAR LOCATION: PCSA IS SECRETED THROUGH THE REGULATED
 CC SECRETORY PATHWAY. PCSB IS A TYPE I MEMBRANE PROTEIN LOCALIZED TO
 CC A PARANUCLEAR POST-GOLGI NETWORK COMPARTMENT IN COMMUNICATION WITH
 CC EARLY ENDOSOMES.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=PCSB; Synonyms=Long;
 CC IsoId=Q04592-1; Sequence=Displayed;
 CC Name=PCSA; Synonyms=Short;
 CC IsoId=Q04592-2; Sequence=VSP_005439;
 CC -!- TISSUE SPECIFICITY: PCSA is expressed in most tissues but is most
 CC abundant in the intestine and adrenal. PCSB is expressed in the
 CC intestine, adrenals and lung but not in the brain.
 CC -!- DEVELOPMENTAL STAGE: Weakly expressed throughout the embryo,
 CC except in the developing nervous system, the ribs and the liver,
 CC but markedly upregulated at discrete sites during development. At
 CC E6.5, prominent expression observed in differentiated decidua. At
 CC E7.5, intense expression in extraembryonic endoderm, amnion and
 CC nascent mesoderm. At E8.5, abundant expression in somites and yolk
 CC sac followed by a confinement to dermamyotome compartment. Between
 CC E9.5 and E11.5, abundant expression in AER (thickened ectodermal
 CC cells of limb buds). At E12.5, expression in the limbs is confined
 CC to the condensing mesenchym surrounding the cartilage. At this
 CC stage, strong expression also detected in vertebral and facial
 CC cartilage primordia and in the muscle of the tongue. At E16.5,
 CC abundant expression in epithelial cells of the intestinal villi.
 CC Isoform A is most abundant at all stages but significant levels of
 CC isoform B occur at E12.5.
 CC -!- DOMAIN: The propeptide domain acts as an intramolecular chaperone
 CC assisting the folding of the zymogen within the endoplasmic
 CC reticulum.
 CC -!- DOMAIN: AC 1 and AC 2 (clusters of acidic amino acids) contain
 CC sorting information. AC 1 directs TGN localization and interacts
 CC with the TGN sorting protein PACS-1.
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC -!- SIMILARITY: Contains 1 homo B/p domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D17583; BAA04507.1; -;
 CC EMBL; D12619; BAA02143.1; -;
 CC EMBL; L14932; AAA74636.1; -;
 CC PIR; A48225; A48225.
 CC PIR; S34583; S34583.
 CC HSSP; Q99405; 1MPT.
 CC MROPS; S08.076; -;
 CC MGD; MG1.97515; Pcsks.
 CC InterPro; IPR006212; Furin repeat.
 CC InterPro; IPR009030; Grow_fac_recep.

DR InterPro; IPR000209; Peptidase S8.
 DR InterPro; IPR002884; Peptidase_S8B.
 DR InterPro; IPR009020; Protease_inhib.
 DR Pfam; PF01483; P_protease; 1.
 DR Pfam; PF00082; Peptidase S8; 1.
 DR PRINTS; PD00723; SUBTILISIN.
 DR ProDom; PD00717; p domain; 1.
 DR SMART; SM00261; FU_22.
 DR PROSITE; PS00136; SUBTILASE ASP; 1.
 DR PROSITE; PS00137; SUBTILASE HIS; 1.
 DR PROSITE; PS00138; SUBTILASE SER; 1.
 KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
 KW Cleavage on pair of basic residues; Repeat; Alternative splicing;
 KW Transmembrane.
 FT SIGNAL 1 34
 FT PROPEP 35 116
 FT CHAIN 117 1877
 FT
 FT DOMAIN 117 1768 PROTEIN CONVERTASE SUBTILISIN/KEXIN
 FT TRANSMEM 1769 1799 TYPE 5.
 FT DOMAIN 1790 1877 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 117 452 POTENTIAL.
 FT DOMAIN 464 602 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 638 1753 CATALYTIC.
 FT DOMAIN 1825 1844 HOMO B.
 FT DOMAIN 1856 1877 CYS-RICH MOTIF (CRM) REGION.
 FT SITE 116 117 AC 1.
 FT SITE 521 523 AC 2.
 FT SITE 173 173 CLEAVAGE (AUTO-) (BY SIMILARITY).
 FT ACT_SITE 214 214 CELL ATTACHMENT SITE (POTENTIAL).
 FT ACT_SITE 388 388 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 227 227 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 383 383 N-LINKED (GLCNAC..) (POTENTIAL).
 FT CARBOHYD 667 667 N-LINKED (GLCNAC..) (POTENTIAL).
 FT CARBOHYD 754 754 N-LINKED (GLCNAC..) (POTENTIAL).
 FT CARBOHYD 804 804 N-LINKED (GLCNAC..) (POTENTIAL).
 FT CARBOHYD 854 854 N-LINKED (GLCNAC..) (POTENTIAL).
 FT CARBOHYD 951 951 N-LINKED (GLCNAC..) (POTENTIAL).
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC..) (POTENTIAL).
 FT CARBOHYD 1220 1220 N-LINKED (GLCNAC..) (POTENTIAL).
 FT CARBOHYD 1317 1317 N-LINKED (GLCNAC..) (POTENTIAL).
 FT CARBOHYD 1523 1523 N-LINKED (GLCNAC..) (POTENTIAL).
 FT CARBOHYD 1711 1711 N-LINKED (GLCNAC..) (POTENTIAL).
 FT CARBOHYD 1733 1733 N-LINKED (GLCNAC..) (POTENTIAL).
 FT VARSPPLIC 878 915 GEYLDQHCQTCESACAKWGPTQEDCISCPVTRVLD ->
 FT ATESWAEGGFCMLVKNNLCQRYLQQLCKCTPFG
 FT (in isoform PCSA).
 FT /FTId=VSP_005438.
 FT /FTId=VSP_005439.
 FT VARSPPLIC 916 1877 Missing (in isoform PCSA).
 FT
 FT SEQUENCE 1877 AA; 209287 MW; EC850E2D20E8A1C3 CRC64;
 Query Match 12.3%; Score 186.5; DB 1; Length 1877;
 Best Local Similarity 25.0%; Pred. No. 6.5e-06;
 Matches 66; Conservative 31; Mismatches 90; Indels 77; Gaps 15;
 QY 39 SQCCQCGCATCSYNG--CLSCXPLFPALERIGMKQIGVCLSCPSGVYGYTPDI--N 94
 Db 1209 NQPCSSCKTC---NGSLCASCPTGMYLWLO-----ACVSPCPOG----TWPSVTSG 1253
 QY 95 KCTKCKADCDTCFNKNPCTKCKS---GFYLHLGKCLDNCPEGLANNHTWE-CVSIVHC 149
 Db 1254 SCCKCEDVCSCGADLCQQLSQPDNTLLHGRGVHSCBPGPYAKGVGVCHECSS--PC 1311
 QY 150 EVSENPMPSCTKKG-----KTC-----GFKRGTRTVRHIIQHPAKGNL-- 190
 Db 1312 KTCENATSCNCSGDFVLHDGVCMKTPKPHAVBGVCKHCPERCQDCIHEKCKEOMP 1371
 QY 191 -----CPPT--NETRKTQVQRKKQKGRGKGRKKKPKNKGSKEAIPDS 236
 Db 1372 DFFLYNDMCHRSCKFPYDMRQCPCHKNC-----LECNPKGDDCKVCADTS 1420
 QY 237 KSLESS---KEIPE----QRENKQ 253

Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
Cleavage on pair of basic residues; Repeat; Alternative splicing;
Transmembrane.

KW	SIGNAL	1	25	POTENTIAL.	
KW	PROPEP	26	110	POTENTIAL.	
FT	CHAIN	111	1696	PROTEIN CONVERTASE SUBTILISIN/KEXIN	
FT				TYPE 5.	
FT	DOMAIN	111	1618	EXTRACELLULAR (POTENTIAL).	
FT	TRANSNEM	1619	1639	POTENTIAL.	
FT	DOMAIN	1640	1696	CYTOPLASMIC (POTENTIAL).	
FT	DOMAIN	111	488	CATALYTIC.	
FT	DOMAIN	496	637	HOMO B.	
FT	DOMAIN	664	1649	CYS-RICH MOTIF (CRM) REGION	
FT	SITE	110	111	CLAVAGE (AUTO-) (BY SIMILARITY).	
FT	ACT_SITE	192	192	CHARGE RELAY SYSTEM (BY SIMILARITY).	
FT	ACT_SITE	233	233	CHARGE RELAY SYSTEM (BY SIMILARITY).	
FT	ACT_SITE	407	407	CHARGE RELAY SYSTEM (BY SIMILARITY).	
FT	CARBOHYD	246	246	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	529	529	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	885	885	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	VARSPLIC	1259	1323	DDTLTDRGECITSCPEGYMDRKKKCYACHPTCKCSDEY	
FT				DDTCTACAGCFLTLDASSCEAGCP -> AENNOQASFCPFA	
FT				PREVSLAEALGHLRYSLTDVPPQSNSPDPTVLGADRRL	
FT				TTATSAGRCA (in isoform C).	
FT				/FTid=VSP_005442.	
FT	VARSPLIC	1324	1696	Missing (in isoform C).	
FT				/FTid=VSP_005443.	
FT	VARSPLIC	1288	1343	CHPTCKCSDEYDITCTACNDVLLTLDASSCEAGCPGQFL	
FT				GHGDCDSCHRECKT -> IARCVDDRDSWCDLAVLRFNC	
FT				VRYFVRCGCTCKLYNEDPRMRGSSQPTQGRN (in	
FT				isoform A).	
FT				/FTid=VSP_005444.	
FT	VARSPLIC	1344	1696	Missing (in isoform A).	
FT				/FTid=VSP_005445.	
FT	SEQUENCE	1696 AA;	188410 MW;	281CBEL784257CBD CRC64;	
QY	Query Match	12.2%;	Score 184.5;	DB 1; Length 1696;	
DB	Best Local Similarity	22.7%;	Pred. No. 8.2e-06;		
QY	Matches	64;	Conservative	25; Mismatches 74; Indels 119; Gaps 13;	
QY	34	MHENVSGGGGATCS--DYNGCLSKCKPLFF-----	64		
DB	1328	LHRGDCDSCHRECKTCDGPHHDNCLSCQPSYLNDQCSTHCPGTEFTYEDSDSGHTVL	1387		
QY	65	-----ALERIGMKQIGVCLSSCPSSGYGYTPDI-NKTKCKA	101		
DB	1388	QCRLCHVNCKTCHGEEDCEMCANDIKYKQDGRCVTECQEGH---YPLDTNECQCWS	1443		
QY	102	DCDTCF--NRNFTCKKSGYLLHKLKDCNCPGL-----EANN	138		
DB	1444	DCETCDGPRNDQCVTCPYNYLLVGLKLEDCPEGYDTMRQKECGCHPSCATCNBGN	1503		
QY	139	HTMECVSIHV-----CEVSEW-----NPMSPTCKKGTCKGKGTETRVREII	181		
DB	1504	Y--NCLSCPYSKLGBCVCPMCEEHYVKEVQICECDNSCKTC---RSTAHDCCLSC	1558		
QY	182	QHP---SAKNLC-----PPTNE-----TRKCTVQRK	205		
DB	1559	EAPGYHAMKHLCTACCEGSPENEYCCICHESTRLCITDRE	1600		

RESULT 3

FUR2	DROME	STANDARD;	PRT;	1679 AA.
ID	FUR2 DROME			
AC	P30432; Q24301;			
DT	01-APR-1993 (Rel. 25, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Furin-like protease 2 precursor (EC 3.4.21.75) (Furin 2).			
GN	FUR2 OR CG18734/CG4235			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			

OC Ephydroidea; Drosophilidae; Drosophila.
 RN NCBI_TaxID=7227;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R, Tuebingen, and Iso-1;
 RX MEDLINE=92381036; PubMed=1512259;
 RA Roebroek A.J.M., Creemers J.W.M., Pauli I.G.L., Kurzik-Dumke U.,
 RA Rentrop M., Gateff E.A.P., Leunissen J.A.M., van de Ven W.J.M.;
 RT "Cloning and functional expression of Dfurin2, a subtilisin-like
 RT protease processing enzyme of Drosophila melanogaster with multiple
 RT repeats of a cysteine motif.";
 RL J. Biol. Chem. 267:17208-17215(1992).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Iso-1;
 RX MEDLINE=95186060; PubMed=7880443;
 RA Roebroek A.J.M., Ayoubi T.A.Y., Creemers J.W.M., Pauli I.G.L.,
 RA van de Ven W.J.M.;
 RT "The Dfur2 gene of Drosophila melanogaster: genetic organization,
 RT expression during embryogenesis, and pro-protein processing activity
 RT of its translational product Dfurin2.";
 RL DNA Cell Biol. 14:223-234(1995).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Holt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hoston D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
 RA Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 CC -!- FUNCTION: Furin is likely to represent the ubiquitous endoprotease
 CC activity within constitutive secretory pathways and capable of
 CC cleavage at the RK(K/R)R consensus motif (By similarity).
 CC
 CC -!- CATALYTIC ACTIVITY: Release of mature proteins from their
 CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa
 CC can be any amino acid and Yaa is Arg or Lys. Releases albumin,
 CC complement component C3 and von Willebrand factor from their
 CC respective precursors.
 CC
 CC -!- TISSUE SPECIFICITY: Transient expression in a subset of central

CC nervous system neurons during embryonic stages 12-13. Expression
 CC in developing tracheal tree from stage 13 to end of embryonic
 CC development.
 CC
 CC -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.
 CC -!- SIMILARITY: Belongs to peptidase family S9. Furin subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC
 DR EMBL; M94375; AAA28551.1; -;
 DR EMBL; L33831; AAA69860.1; -;
 DR EMBL; AE003502; AAF48598.1; -;
 DR PIR; A43434; A43434.
 DR HSSP; Q99405; LMPT.
 DR MEROPS; S08_049; -;
 DR FlyBase; FBgn0004598; Fur2.
 DR GO; GO:0004276; F: furin activity; IDA.
 DR InterPro; IPR006212; Furin_repeat.
 DR InterPro; IPR009030; Grow_fac_recep.
 DR InterPro; IPR00209; Peptidase_S8.
 DR InterPro; IPR002884; Peptidase_S8B.
 DR InterPro; IPR009020; Protease_inhib.
 DR Pfam; PF01483; P_protein; 2.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR ProDom; PD000717; P_domain; 1.
 DR SMART; SM00261; FU_10.
 DR PROSITE; PS00136; SUBTILASE ASP; 1.
 DR PROSITE; PS00137; SUBTILASE HIS; 1.
 DR PROSITE; PS00138; SUBTILASE SER; 1.
 KW Hydrolase; Serine protease; Glycoprotein; Signal; Transmembrane;
 KW Multigene family; Zymogen; Repeat.
 FT SIGNAL; 1; ? POTENTIAL.
 FT PROPEP; 318 POTENTIAL.
 FT CHAIN; 319 1679 FURIN-LIKE PROTEASE 2.
 FT ACT_SITE; 417 417 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE; 456 456 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE; 637 637 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DOMAIN; 961 1443 10 X TANDEM REPEATS, CYS-RICH.
 FT REPEAT; 961 1006 1.
 FT REPEAT; 1007 1056 2.
 FT REPEAT; 1057 1103 3.
 FT REPEAT; 1104 1152 4.
 FT REPEAT; 1153 1204 5.
 FT REPEAT; 1205 1253 6.
 FT REPEAT; 1254 1298 7.
 FT REPEAT; 1299 1345 8.
 FT REPEAT; 1346 1392 9.
 FT REPEAT; 1393 1443 10.
 FT TRANSMEM; 1512 1532 POTENTIAL.
 FT DOMAIN; 1533 1679 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD; 3 3 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD; 109 109 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD; 130 130 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD; 205 205 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD; 442 442 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD; 480 480 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD; 927 927 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD; 1060 1060 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD; 1181 1181 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD; 1274 1274 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD; 1277 1277 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD; 1439 1439 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CONFLICT; 152 153 MISSING (IN REF. 1).
 FT CONFLICT; 177 177 V -> F (IN REF. 1).
 FT CONFLICT; 213 213 V -> VDQL (IN REF. 1).
 SQ SEQUENCE 1679 AA; 183369 MW; 3F9E749F0B021CF6 CRC64;

```
Query Match 11.68; Score 176.5; DB 1; Length 1679;
Best Local Similarity 28.08; Pred.No. 3e-05; Gaps 12;
Matches 60; Conservative 24; Mismatches 77; Indels 53;

QY 3 HLRLISWLPFIILNFMIEYIGSQNASRGRORRHPNVSGQGGCATCSY-NGCLSCKPR 61
DB 1037 HLHVID-LAVCLQCPDGVFENS-----RNRTCPV-----CEPNCASCODHPEYCTSCDHH 1086

QY 62 LPFALERIGMKQGVCLSSCPGSGYGYTRYDPDINKCTKCKADCDTCP--NNPCTKCSGF 119
DB 1087 LVNHEHK-----CYSACPLDYET---BDNKCAPCHSTCATCNGPTDQDCITCRSSR 1135

QY 120 YLHLGKCLDNCPEGLRANNHMECVSVHCVSEWNPMSPTCKGKTCGFKRGYETEV-R 178
DB 1136 YAWQNKRLSCPDPGYADKKRLECM-----PQRCCKTC-----TSNGVCS 1176

QY 179 EIIQHPSAKGNLCPPTNTRKCTVQRKK-CQKGE 211
DB 1177 ECLQNWIT-----LNKRDKCVSGSGGCSSEB 1202

RESULT 4
ID PAC4 HUMAN STANDARD; PRT: 969 AA.
AC P29132; Q15099; Q15100; Q9UEG7; Q9UEJ1; Q9UEJ2; Q9UEJ7; Q9UEJ8;
AC Q9UEJ9; Q9Y4G9; Q9Y4H0; Q9Y4H1;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Paired basic amino acid cleaving enzyme 4 precursor (EC 3.4.21.-)
DE (Subtilisin/kexin-like protease PAC4) (Subtilisin-like proprotein
DE convertase 4) (SPC4).
OS PAC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A. (ISOFORMS PAC4A-I AND PAC4B).
RC TISSUE=Hepatoma, and Kidney;
RX MEDLINE=92075167; PubMed=1741956;
RA Kiefer M.C., Tucker J.B., Joh R., Landeberg K.B., Saltman D.,
RA Barr P.J.;
RA "Identification of a second human subtilisin-like protease gene in
RT the tes/fps region of chromosome 15";
RT DNA Cell Biol. 10:757-769(1991).
[2]
RN SEQUENCE FROM N.A. (ISOFORMS PAC4A-I AND PAC4D).
RC TISSUE=Placenta;
RX MEDLINE=94235049; PubMed=8179631;
RA Teuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,
RA Matsuda Y.;
RA "Identification of novel cDNAs encoding human kexin-like protease,
RT PAC4A isoforms.";
RT Biochem. Biophys. Res. Commun. 200:943-950(1994).
[3]
RN ERRATUM.
RP MEDLINE=95071480; PubMed=7980617;
RA Teuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,
RA Matsuda Y.;
RT "Identification of novel cDNAs encoding human kexin-like protease,
RT PAC4A isoforms.";
RL Biochem. Biophys. Res. Commun. 204:1381-1382(1994).
[4]
RN SEQUENCE FROM N.A. (ISOFORM PAC4A-II).
RC TISSUE=Placenta;
RA Mori K., Imamaki A., Kii S., Nagamune H., Nagahama M., Teuji A.,
RA Matsuda Y.;
RT "Identification of a novel PAC4A isoform, PAC4A-II";
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
[5]
RN SEQUENCE FROM N.A. (ISOFORMS PAC4A-I AND PAC4B-II).
RC TISSUE=Cerebellum;
```

```
RX MEDLINE=97335942; PubMed=9192737;
RA Mori K., Kii S., Teuji A., Nagahama M., Imamaki A., Hayashi K.,
RA Akamatsu T., Nagamune H., Matsuda Y.;
RT "A novel human PAC4A isoform, PAC4B is an active processing protease
RL containing a hydrophobic cluster at the carboxy terminus.";
RL J. Biochem. 121:941-948(1997).
[6]
RN SEQUENCE FROM N.A. (ISOFORMS PAC4A-I; PAC4A-II; PAC4A-III; PAC4A-IV; PAC4A-V;
RP PAC4A-I AND PAC4A-II).
RX MEDLINE=98021085; PubMed=9378725;
RA Teuji A., Hine C., Tamai Y., Yonemoto K., Mori K., Yoshida S.,
RA Bando M., Sakai E., Mori K., Akamatsu T., Matsuda Y.;
RT "Genomic organization and alternative splicing of human PAC4A (SPC4),
RT kexin-like processing endoprotease.";
RL J. Biochem. 122:438-452(1997).
[7]
RN ALTERNATIVE SPLICING (ISOFORM PAC4CS).
RX MEDLINE=97064242; PubMed=8906861;
RA Zhong M., Benjannet S., Lazure C., Munzer S., Seidah N.G.;
RT "Functional analysis of human PAC4A-A and PAC4A-C isoforms:
RT identification of a new PAC4A-CS isoform.";
RL FEBS Lett. 396:31-36(1996).
[8]
RN CHARACTERIZATION.
RX MEDLINE=99233559; PubMed=10215603;
RA Sucic J.P., Moehring J.M., Inocencio N.M., Luchini J.W.,
RA Moehring T.J.;
RT "Endoprotease PAC4A is Ca2+-dependent and temperature-sensitive and
RT can partly rescue the phenotype of a furin-deficient cell strain.";
RL Biochem. J. 339:639-647(1999).
[9]
RN PROCESSING.
RX MEDLINE=96408849; PubMed=9738469;
RA Nagahama M., Taniguchi T., Hashimoto E., Imamaki A., Mori K.,
RA Teuji A., Matsuda Y.;
RT "Biosynthetic processing and quaternary interactions of proprotein
RT convertase SPC4 (PAC4A)".
RL FEBS Lett. 434:155-159(1998).
-!- FUNCTION: Likely to represent an endoprotease activity within the
CC constitutive secretory pathway, with unique restricted
CC distribution in both neuroendocrine and non-neuroendocrine tissues
CC and capable of cleavage at the RX(K/R)R consensus motif.
CC -!- CATALYTIC ACTIVITY: Release of mature proteins from their
CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa
CC can be any amino acid and Yaa is Arg or Lys.
CC -!- COFACTOR: Calcium (potential).
CC -!- SUBUNIT: The PAC4A-I precursor protein seems to exist in the
CC reticulum endoplasmic as both a monomer and a dimer-sized complex
CC whereas mature PAC4A-I exists only as a monomer, suggesting that
CC propeptide cleavage affects its tertiary or quaternary structure.
CC and PAC4A-III are not secreted and remain probably in zymogen form
CC in endoplasmic reticulum. PAC4A-I and PAC4A-II are retained
CC intracellularly probably through a hydrophobic cluster in their C-
CC terminus. PAC4B might be secreted.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=8;
CC Name=PAC4A-I; Synonyms=PAC4A;
CC IsoId=P29122-1; Sequence=Displayed;
CC Name=PAC4A-II;
CC IsoId=P29122-2; Sequence=VSP_005436;
CC Name=PAC4A-III;
CC IsoId=P29122-3; Sequence=VSP_005428, VSP_005429;
CC Note=Probably enzymatically inactive;
CC Name=PAC4A-IV;
CC IsoId=P29122-4; Sequence=VSP_005432, VSP_005433;
CC Note=Probably enzymatically inactive;
CC Name=PAC4A-V;
CC IsoId=P29122-5; Sequence=VSP_005430, VSP_005431;
CC Note=Probably enzymatically inactive;
CC Name=PAC4A-IV;
CC IsoId=P29122-6; Sequence=VSP_005427, VSP_005434, VSP_005435;
CC Note=Probably enzymatically inactive;
```


CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa
 CC can be any amino acid and Yaa is Arg or Lys.
 CC -|- SUBCELLULAR LOCATION: PCSA IS SECRETED THROUGH THE REGULATED
 CC SECRETORY PATHWAY. PCSB IS A TYPE I MEMBRANE PROTEIN LOCALIZED TO
 CC A PARANUCLEAR POST-GOLGI NETWORK COMPARTMENT IN COMMUNICATION WITH
 CC EARLY ENDOSOMES (BY SIMILARITY).
 CC -|- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=PCSB; Synonyms=Long;
 CC IsoId=P41413-1; Sequence=Displayed;
 CC Name=PCSA; Synonyms=Short;
 CC IsoId=P41413-2; Sequence=VSP_005440, VSP_005441;
 CC -|- TISSUE SPECIFICITY: Expressed in the intestine, brain, adrenal
 CC gland, anterior pituitary, thyroid, ovaries, testis and lung.
 CC Highest levels are found in the gut, duodenum, jejunum and ileum.
 CC Expression is higher in female than in male reproductive organs.
 CC -|- DEVELOPMENTAL STAGE: First detected at E9 in highly restricted
 CC regions of the neural tube, in caudal myotomes, and at the
 CC materno-embryonic junction of the uterus. At E10, restricted
 CC expression is detected in the optic and otic vesicles, the roof of
 CC midbrain, and trunk myotomes. By midgestation (E13-E16),
 CC expression in the developing nervous system has expanded to
 CC multiple regions including hippocampus, thalamus, hypothalamus,
 CC brain stem, and spinal cord. Expression is also detected in
 CC several peripheral organ systems, including gut, lung, adrenal and
 CC kidney primordia.
 CC -|- DOMAIN: The propeptide domain acts as an intramolecular chaperone
 CC assisting the folding of the zymogen within the endoplasmic
 CC reticulum.
 CC -|- DOMAIN: AC 1 and AC 2 (clusters of acidic amino acids) contain
 CC sorting information. AC 1 directs TGN localization and interacts
 CC with the TGN sorting protein PACS-1.
 CC -|- SIMILARITY: Belongs to peptidase family S8.
 CC -|- SIMILARITY: Contains 1 homo B/P domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; LI4933; AAA99906.1; --
 CC EMBL; U47014; AAA87888.1; --
 CC PIR; B48225; B48225.
 CC HSP; Q99405; 1MPT.
 CC MEROPS; S08.076; --
 CC InterPro; IPR006212; Purin repeat.
 CC InterPro; IPR002029; Peptidase_S8.
 CC InterPro; IPR002884; Peptidase_S8B.
 CC InterPro; IPR009020; Protease_inhib.
 CC Pfam; PF01483; P_protein; 1.
 CC PRINTS; PR00723; SUBTILISIN.
 CC ProDom; PD000717; P_domain; 1.
 CC SMART; SM00261; FU; 6.
 CC PROSITE; PS00136; SUBTILASE ASP; 1.
 CC PROSITE; PS00137; SUBTILASE HIS; 1.
 CC PROSITE; PS00138; SUBTILASE SER; 1.
 CC Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
 CC Cleavage on pair of basic residues; Alternative splicing; Repeat;
 CC Transmembrane.
 CC SIGNAL 1 34 BY SIMILARITY.
 CC PROPEP 35 116 BY SIMILARITY.
 CC CHAIN 117 1877 PROTEIN CONVERTASE SUBTILISIN/KEXIN
 CC TYPE 5.
 CC DOMAIN 117 1768 EXTRACELLULAR (POTENTIAL).
 CC TRANSMEM 1769 1789 POTENTIAL.
 CC DOMAIN 1790 1877 CYTOPLASMIC (POTENTIAL).
 CC DOMAIN 117 452 CATALYTIC.
 CC DOMAIN 464 602 HOMO B.
 CC DOMAIN 638 1753 CYS-RICH MOTIF (CRM) REGION.

FT	DOMAIN	1825	1844	AC 1.
FT	DOMAIN	1856	1877	AC 2.
FT	SITE	116	117	CLEAVAGE (AUTO-) (BY SIMILARITY).
FT	SITE	521	523	CELL ATTACHMENT SITE (POTENTIAL).
FT	ACT SITE	173	173	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT SITE	214	214	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT SITE	388	388	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	CARBOHYD	227	227	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	383	383	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	667	667	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	754	754	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	804	804	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	854	854	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1710	1710	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1732	1732	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	878	915	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX -->
FT				ATEESWAEGGFCMLVKKNNLCQKRVLOQLCKCTCTFOG
FT				(in isoform PCSA).
FT	VARSPLIC	916	1877	Missing (in isoform PCSA).
FT				/FTID=VSP_005440.
FT	SEQUENCE	1877 AA;	207888 MW;	890955DC60534444 CRC64;
Query Match		11.1%;	Score 168;	DB 1; Length 1877;
Best Local Similarity		26.3%;	Pred. No. 0.00014;	
Matches		49;	Conservative	25; Mismatches 54; Indels 58; Gaps 11;
Qy	36	PVYSQ-GCOG----	GCATCSYNGCLSKPLPALERIGKQIGVCLSSCPGSGYGYTRY	90
Db	640	PECSEVGCDGPDHCTDCLHYHVKNNTR-----	ICVSSCPGHP----	H 682
Qy	91	PDINKTKKADCDTCENKFP--CTKCKSGFYHL--LGKCLDNCPEGLANNHTMECVSI	146	
Db	683	ADKCRKCKAPNCESCFCGSHADQCLSKYGYFLNEETSSCVAQCPGSGYQDKNNIC----	739	
Qy	147	VHCEVSEWNPSPCTKKGKTC-GFERCTETVRREIIQHPSAKGNLCPTNETRCKTVQK	205	
Db	740	-----GKSENCKTCTGFHNCTE-----	CKGSL---	SLGGRCSV--- 771
Qy	206	KCQKGE 211		
Db	772	TCBDGQ 777		
RESULT 6				
PCSKS HUMAN				
ID	PCSKS HUMAN	STANDARD;	PRT;	913 AA.
AC	Q92824; Q13527;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DB	Protein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.--)			
DB	(Protein convertase PCS5) (Subtilisin/kexin-like protease PCS) (PC6)			
DB	(hPC6)			
GN	PCSK5 OR PCS5 OR PC6.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Theria; Primates; Catarrhini; Hominoidea; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=T-cell;			
RC	MEDLINE=9635380; PubMed=8755538;			
RA	Miranda L., Wolf J., Pichuanes S., Duke R., Franzusoff A.;			
RT	"Isolation of the human PC6 gene encoding the putative host protease			
RT	for HIV-1 gp160 processing in CD4+ T lymphocytes.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 93:7695-7700(1996).			
RN	[2]			
RP	REVISIONS.			
RA	Franzusoff A., Miranda L., Wolf J., Pichuanes S., Lu Y., Duke R.;			
RL	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 15-913 FROM N.A.			

RA Reudelhuber T.L.;
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Likely to represent a widespread endoprotease activity
 CC within the constitutive and regulated secretory pathway. Capable
 CC of cleavage at the RX(K/R)R consensus motif.
 CC -!- CATALYTIC ACTIVITY: Release of mature proteins from their
 CC propeptides by cleavage of Arg-Xaa-Yaa-ARG-|-Zaa bonds, where Xaa
 CC can be any amino acid and Yaa is Arg or Lys.
 CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=1;
 CC Comment=2 isoforms may be produced;
 CC Name=1;
 CC IsoId=Q92824-1; Sequence=Displayed;
 CC TISSUE SPECIFICITY: Expressed in T-lymphocytes.
 CC -!- DOMAIN: The propeptide domain acts as an intramolecular chaperone
 CC assisting the folding of the zymogen within the endoplasmic
 CC reticulum.
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC -!- SIMILARITY: Contains 1 homo B/P domain.
 CC -!- SIMILARITY: Contains 1 PLAC domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U56387; AAC50643.2; --
 CC EMBL; U49114; AAA91807.1; --
 CC PIR; G02428; G02428.
 CC HSP; Q99405; 1MPT.
 CC MEROPS; S08.076; --
 CC Genew; HGNC:8747; PCSK5.
 CC MIM; 600488; --
 CC GO; GO:0005615; C:extracellular space; TAS.
 CC GO; GO:0004252; F:serine-type endopeptidase activity; TAS.
 CC GO; GO:0007267; P:cell-cell signaling; TAS.
 CC GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
 CC InterPro; IPR006212; Funin repeat.
 CC InterPro; IPR000209; Peptidase S8.
 CC InterPro; IPR002884; Peptidase_S8B.
 CC InterPro; IPR009020; Protease_inhib.
 CC Pfam; PF01483; P:protease; 1.
 CC Pfam; PF00082; Peptidase S8; 1.
 CC PRINTS; PR00723; SUBTILISIN.
 CC ProDom; PD000717; P domain; 1.
 CC SMART; SM00261; FU_5.
 CC PROSITE; PS50900; PLAC; 1.
 CC PROSITE; PS00136; SUBTILASE ASP; 1.
 CC PROSITE; PS00137; SUBTILASE HIS; 1.
 CC PROSITE; PS00138; SUBTILASE_SER; 1.
 CC Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
 KW Cleavage on pair of basic residues; Repeat; Alternative splicing.
 FT SIGNAL 1 32
 FT PROPEP 33 114
 FT CHAIN 115 913
 FT TYPE 5.
 FT DOMAIN 115 454
 FT DOMAIN 462 600
 FT DOMAIN 636 868
 FT DOMAIN 869 913
 FT SITE 114 115
 FT SITE 519 521
 FT ACT_SITE 171 171
 FT ACT_SITE 212 212
 FT ACT_SITE 386 386
 FT ACT_SITE 225 225
 FT CARBOHYD 381 381
 FT CARBOHYD 665 665
 FT CARBOHYD 752 752
 FT CARBOHYD 752 752

FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 852 852 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 118 118 S -> F (IN REF. 3).
 FT CONFLICT 121 121 V -> A (IN REF. 3).
 FT CONFLICT 511 511 R -> A (IN REF. 3).
 FT CONFLICT 601 601 R -> Q (IN REF. 3).
 SQ SEQUENCE 913 AA; 101775 MW; 21389264CAD7546C CRC64;
 Query Match 10.7%; Score 162; DB 1; Length 913;
 Best Local Similarity 22.3%; Pred. No. 0.00018;
 Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;
 QY 21 GSONASRGRRQRMRHENVSGCGGCATCSDYNG--CLSCPKRLPALERIGMKQIGVCL 78
 DB 677 GHYHAK-KRCRKAPN-----CSCPGSGHGDQMSCKYGYFL-----NBEITNSCV 721
 QY 79 SSCPSG-YGTTPYDINKCTCKACDCTCFNKNFCTKCKSGPYLHLGLCLNDNPEGLEAN 137
 DB 722 THCPDGSYQDTKK---NLCKKSENCKTCTEFHNCCTCRDGLSLQGRCSVSCDGRYFN 778
 QY 138 NHTMR-----CVS-----IVHCEVSEW-----NPWSPCTKK 163
 DB 779 GDCQCPCHRFCAFCAGAGADGGINCTGYFMEDGRCVQSCSIYFYDHSSENGYKCKKC 838
 QY 164 GKTC-----GPKRGTE-----TRVEIIQHPSAKGNLCPTTNETRKTCTV 202
 DB 839 DRISLTCNGPGFNCCTSCPSGYLLDGMQMGAIKDAEESWARGGFQMLVKXNULC-- 896
 QY 203 QRKCKQK 209
 DB 897 QRKVLQK 903
 RESULT 7
 ID WIF1 HUMAN STANDARD; PRT; 379 AA.
 AC Q9Y5W5; Q8WVG4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Wnt inhibitory factor 1 precursor (WIF-1).
 GN WIF1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99215557; PubMed=10201374;
 RA Heich J.-C., Kodjabachian L., Rebber M.L., Rattner A.,
 RA Smallwood P.M., Samos C.H., Nusse R., Dawid I.B., Nathans J.;
 RA "A new secreted protein that binds to Wnt proteins and inhibits their
 RA activities.";
 RL Nature 398:431-436(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Krausberg R.D., Collins F.S., Wagner L., Shennen C.M., Schuler G.D.,
 RA Altschul S.P., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh P.,
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalhus D.B.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).
 CC -!- FUNCTION: Binds to Wnt proteins and inhibits their activities. May
 be involved in mesoderm segmentation.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Contains 5 EGF-like domains.
 CC -!- SIMILARITY: Contains 1 WIF domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF122922; AAD25402.1; --
 DR EMBL; BC018037; AAH18037.1; --
 DR PIR; A59180; A59180.
 DR HSP; P00743; LCCP.
 DR Genew; HGNC:18081; WIF1.
 DR MIM; 605186; --
 DR InterPro; IPR006209; EGF like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR003306; WIF.
 DR Pfam; PF00008; EGF; 5.
 DR Pfam; PF02019; WIF; 1.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00181; EGF; 5.
 DR SMART; SM00469; WIF; 1.
 DR PROSITE; PS00022; EGF_1; 5.
 DR PROSITE; PS01186; EGF_2; 4.
 DR PROSITE; PS00026; EGF_3; 5.
 DR PROSITE; PS00814; WIF; 1.
 DR Repeat; EGF-like domain; Signal; Developmental protein;
 KW Wnt signaling pathway.
 FT SIGNAL 1 28
 FT CHAIN 29 379
 FT DOMAIN 38 176
 FT DOMAIN 177 208
 FT DOMAIN 209 240
 FT DOMAIN 241 272
 FT DOMAIN 273 304
 FT DOMAIN 305 336
 FT DISULFID 177 186
 FT DISULFID 182 192
 FT DISULFID 198 200
 FT DISULFID 209 218
 FT DISULFID 214 224
 FT DISULFID 230 232
 FT DISULFID 241 250
 FT DISULFID 246 256
 FT DISULFID 262 264
 FT DISULFID 273 282
 FT DISULFID 278 288
 FT DISULFID 294 296
 FT DISULFID 305 314
 FT DISULFID 310 320
 FT DISULFID 326 328
 FT CARBOHYD 88 88
 FT CARBOHYD 245 245
 FT CONFLICT 166 166
 FT CONFLICT 178 178
 FT CONFLICT 178 178
 SQ SEQUENCE 379 AA; 41528 MW; 32EC54D60529EP96 CRC64;
 Query Match 10.5%; Score 158.5; DB 1; Length 379;
 Best Local Similarity 24.9%; Pred. No. 0.00014;
 Matches 60; Conservative 26; Mismatches 72; Indels 83; Gaps 16;

QY 42 CCGCAT---CSDYNGC-----LSCPRLPFALEIRGNKQIGVCLSS---C 81
 DB 182 CPGCRNGGFCNERRICEPDGPHGPHCEKALCTPONG-----GLCVTPGFCIC 232
 QY 82 PSYGYTRYDINKTKC-KADCD--TCFNKPNCTCKSGFYHLGKCLDNCPEGLRANH 139
 DB 233 PPGFVG-----VNCDKANCSITCFNGGTC-----FY--PGKCI--CPGLEGR-- 271
 QY 140 TMECVSIVHCEVSEWNPSPCTKKGKGTGPKRGTTETRVREIIQHPSAKGNLCPP----- 193
 DB 272 -----QCHSKCP--QPCRNKGKCTG--KSKCKSGYGGDLCSKVPCEPGCGANG 318
 QY 194 -TNETRTKTVQKKCKQKGERGKGRERK-----KFNKGSKEAIPDSKLSSESKTIP 246
 DB 319 TCHEPNKC-----CQEGHGRHCNKRVEASLIHALRPAGALRQHTPSLKAERDRDP 373
 QY 247 E 247
 DB 374 E 374

RESULT 8
 WIF1_XENLA
 ID WIF1_XENLA STANDARD; PRT; 374 AA.
 AC Q9W6F8;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DB WNT inhibitory factor 1 precursor (WIF-1).
 GN WIF1.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
 OC Xenopodidae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99215557; PubMed=10201374;
 RA Hsieh J.-C., Kodjabachian L., Rebert M.L., Rattner A.,
 RA Smallwood P.M., Samos C.H., Nusse R., David I.B., Nathans J.;
 RA "A new secreted protein that binds to Wnt proteins and inhibits their
 activities.";
 RT Nature 398:431-436(1999).
 RL Nature 398:431-436(1999).
 CC -!- FUNCTION: Binds to Wnt proteins and inhibits their activities. May
 be involved in mesoderm segmentation.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: During somatogenesis, expressed predominantly
 in unsegmented paraxial presomitic mesoderm and, to a much lesser
 extent, in newly segmented somites.
 CC -!- DEVELOPMENTAL STAGE: First expressed at neurula stages.
 CC -!- SIMILARITY: Contains 5 EGF-like domains.
 CC -!- SIMILARITY: Contains 1 WIF domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF122924; AAD25404.1; --
 DR HSP; P00740; IEDM.
 DR InterPro; IPR006209; EGF like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR003306; WIF.
 DR Pfam; PF00008; EGF; 5.
 DR Pfam; PF02019; WIF; 1.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00181; EGF; 4.
 DR SMART; SM00469; WIF; 1.
 DR PROSITE; PS00022; EGF_1; 5.

```

DR PROSITE; PS01186; EGF 2; 5.
DR PROSITE; PS00226; EGF_3; 4.
DR PROSITE; PS0814; WIF; 1.
KW Repeat; EGF-like domain; Signal; Developmental protein;
Wnt signaling pathway.
FT SIGNAL 1 28
FT CHAIN 29 374
FT DOMAIN 33 171
FT DOMAIN 172 203
FT DOMAIN 204 235
FT DOMAIN 236 267
FT DOMAIN 268 299
FT DOMAIN 300 331
FT DISULFID 172 181
FT DISULFID 177 187
FT DISULFID 193 195
FT DISULFID 204 213
FT DISULFID 209 219
FT DISULFID 225 227
FT DISULFID 236 245
FT DISULFID 241 251
FT DISULFID 257 259
FT DISULFID 268 277
FT DISULFID 273 283
FT DISULFID 289 291
FT DISULFID 300 309
FT DISULFID 305 315
FT DISULFID 321 323
FT CARBOHYD 83 83
FT CARBOHYD 240 240
SQ SEQUENCE 374 AA; 41071 MW; E26F973B0F00ACF8 CRC64;

Query Match 10.4%; Score 157.5; DB 1; Length 374;
Best Local Similarity 24.3%; Pred. No. 0.00017;
Matches 57; Conservative 22; Mismatches 85; Indels 71; Gaps 12;

QY 42 CQGGCAT---CSDYNGC-----LSCKPRLFPALERIGMKQIGVC-----LSSC 81
DB 177 CTGCGRNGFGFCDRHVCECDPGFGPHCEKALCPKRWG-----GLCTPGLCIC 227
QY 82 PSQYGYTRYPDINKTKCKADCDTCFNKNFCTKSGFYHLGLKCLDNCPEGLEANNHTM 141
DB 228 PPGYGINCDKVNCTTHC-LNGGTCP-----YPGKCI--CPSGYEGE----- 266
QY 142 ECYSIVHCEVSEWNPSPCKTKGCTGPKGCTHTVRELIHPSAKGNLCPTNETKCT 201
DB 267 -----QCETSKQ--QPCRNGGCKSG---KNKCKSGYQGDLCSPKVPCEPCGAHGT 315
QY 202 VQRKCKQKGRGKGRERKPK-----KPNKGSKEAIPDSKLSSESKPE 247
DB 316 IEPNKCQCKE-GWNGRYCNKKYGSNLNMLNLRPTGSRNRHTPSPKRTEDQALPE 369

RESULT 9
PAC4 RAT STANDARD; PRT; 937 AA.
AC Q63415;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Paired basic amino acid cleaving enzyme 4 precursor (BC 3.4.21.-)
DE (Subtilisin/kexin-like cleaving enzyme 4 precursor)
DE (Subtilisin/kexin-like protease PACB4) (Subtilisin-like proprotein
DE convertase 4) (SPC4).
GN PACB4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Hypothalamus, and Pituitary;
RX MEDLINE=94349873; PubMed=8070361;
RA Johnson R.C., Darlington D.N., Hand T.A., Bloomquist B.T., Mains R.E.;

```

```

RT "PACB4: a subtilisin-like endoprotease prevalent in the anterior
RT pituitary and regulated by thyroid status.";
RL Endocrinology 135:1178-1185(1994).
CC -!- FUNCTION: Likely to represent an endoprotease activity within the
CC constitutive secretory pathway, with unique restricted
CC distribution in both neuroendocrine and non-neuroendocrine tissues
CC and capable of cleavage at the RX(K/R)R consensus motif.
CC -!- CATALYTIC ACTIVITY: Release of mature proteins from their
CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa
CC can be any amino acid and Yaa is Arg or Lys.
CC -!- COFACTOR: Calcium (Potential).
CC -!- TISSUE SPECIFICITY: High expression in the anterior pituitary and
CC in several brain regions, the atrium, and the ventricle.
CC -!- DOMAIN: The propeptide domain acts as an intramolecular chaperone
CC assisting the folding of the zymogen within the endoplasmic
CC reticulum.
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC -!- SIMILARITY: Contains 1 homo B/P domain.
CC -!- SIMILARITY: Contains 1 PLAC domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
EMBL; L31894; AAA61987.1; -.
DR F1R; I53282; I53282.
DR HSP; Q99405; LMPT.
DR MEROPS; S08.075; -.
DR InterPro; IPR006212; Purin repeat.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR002884; Peptidase_S8B.
DR InterPro; IPR009020; Protease_inhib.
DR Pfam; PF01483; P_protein; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR ProDom; PD000717; P_domain; 1.
DR SMART; SM00261; FU; 5.
DR PROSITE; PS0900; PLAC; 1.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal; Calcium;
KW Cleavage on pair of basic residues; Repeat.
FT SIGNAL 1 45
FT PROPEP 46 132
FT CHAIN 133 937
FT DOMAIN 133 454
FT DOMAIN 477 615
FT DOMAIN 680 898
FT DOMAIN 899 937
FT SITE 132 133
FT SITE 534 536
FT ACT_SITE 186 186
FT ACT_SITE 227 227
FT ACT_SITE 401 401
FT CARBOHYD 240 240
FT CARBOHYD 882 882
FT CARBOHYD 900 900
FT SEQUENCE 937 AA; 104053 MW; P3865557C33705C8 CRC64;

Query Match 10.3%; Score 156.5; DB 1; Length 937;
Best Local Similarity 23.6%; Pred. No. 0.00046;
Matches 51; Conservative 24; Mismatches 80; Indels 61; Gaps 11;

QY 19 YIGSONASRGRRQRMRHFNPNVSGCGCATCSQDN--GCLSKPRLFPALERIGMKQIGV 76
DB 705 YFGDTAARRCR-----CHKGCTCTGKSPQCLSCR-RGFI-----HQETNT 747

```


QY 77 CLSSCPGYYGTRYPDINKCTKCKADCTCFNK-NFCTKCKSGFYHLGKCLDNCPEGLE 135
 Db 748 CVTLCPAGLYADSSQRL--CLRCHPSCKQCVDBPEKSTVCKEGFSLARGSCIPDCBPGTY 805
 QY 136 ANNHTMEC-----VSIHVCEVS-----EWNPSVCTKKGKTCGFKRGTETR 176
 Db 806 FDSSELIRGCEHCTCTCVGSPRECHCAKSHFQDMKCVPAQGE-----GP----- 853
 QY 177 VREIIOHPSAKGNLCPTNTRKCTVQRKQKQGER 212
 Db 854 -----YPEMPGL--PHKVCRRCDNCLSCGSSR 881

RESULT 10
 FSPO_XENLA STANDARD; PRT; 803 AA.
 ID FSPO_XENLA AC P3547;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE F-spontin precursor.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=93376785; PubMed=8367492;
 RA Ruiz A, Altaba A., Cox C., Jessell T.M., Klar A.;
 RT "Ectopic neural expression of a floor plate marker in frog embryos
 RT injected with the midline transcription factor Pintallavis.";
 RT Proc. Natl. Acad. Sci. U.S.A. 90:8268-8272(1993).
 CC -!- FUNCTION: Promotes the attachment of spinal cord and sensory
 CC neuron cells and the outgrowth of neurites in vitro. May
 CC contribute to the growth and guidance of axons in both the spinal
 CC cord and the PNS.
 CC SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: Expressed at high levels in the floor plate.
 CC -!- SIMILARITY: Contains 6 TSP type-1 domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L09123; AAA19105.1; -;
 DR PIR; A47723; A47723.
 DR InterPro; IPR002861; Reeler.
 DR InterPro; IPR000884; TSP1.
 DR Pfam; PF02014; Reeler; 1.
 DR Pfam; PF00090; tsp_1; 6.
 DR SMART; SM00209; TSP1; 6.
 DR PROSITE; PS50092; TSP1; 6.
 KW Glycoprotein; Signal; Repeat; Cell adhesion.
 FT SIGNAL 1 23
 FT CHAIN 24 803
 FT DOMAIN 436 489
 FT DOMAIN 495 549
 FT DOMAIN 552 605
 FT DOMAIN 608 662
 FT DOMAIN 664 717
 FT DOMAIN 750 802
 FT CARBOHYD 210 210
 FT CARBOHYD 677 677
 SQ SEQUENCE 803 AA; 90702 MW; D3A54E329548AED9 CRC64;

Query Match 10.3%; Score 156; DB 1; Length 803;
 Best Local Similarity 24.8%; Pred. No. 0.00044;

Matches 58; Conservative 35; Mismatches 91; Indels 52; Gaps 12;
 QY 38 VSQCGCGCATCDYNGCLSKPRLPALRIGMKQIGVCLSSCPSS--GYGTRYPDINK 95
 Db 544 VNECEPSSCIVTEWAEWBECS-----ATCRMGKKRHRMIKMTPADGSMCKADTTVEVK 598
 QY 96 C--TKCK-----ADCTCFNKNFCTKCKSGFYHLGKCLDNCPEGLEANN--- 138
 Db 599 CMPECHTIPCVLSPMSEWSDCVTCGKTRTRQR---MLKSPSELGDCNBELELKQVEK 655
 QY 139 -HMECVSIVHCEVSENNPSPCTKCKTCTGKGTETRTVREIIQHPSAKGNLCPTNTR 197
 Db 656 CMLPECP--ISCLTETSYWSEC---NKSCG--KGMIRMTMTWEPQFGGAVCPETVQR 708
 QY 198 RKCTVQRKCKQGERGKGRKRKPKNKGESKEAIPDSKLSKSKSIPRENNKQ 253
 Db 709 KKCL--RKCK-----SSGNERRHLK-----DAREKRSEKIKEDSDGQ 747

RESULT 11
 NTC2_MOUSE STANDARD; PRT; 2470 AA.
 ID NTC2_MOUSE AC Q35516; Q06008; Q60941;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Notch
 DE B).
 GN NOTCH2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Thymus;
 RA Hamada Y., Higuchi M., Tsujimoto Y.;
 RT "Complete amino acid sequence and multifunctional transcripts encoded by a
 RT single copy of mouse Notch2 gene.";
 RT Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 316-1518 FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
 RX MEDLINE=93178563; PubMed=8440332;
 RA Lardelli M., Lendahl U.;
 RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a
 RT wide variety of tissues.";
 RL Exp. Cell Res. 204:364-372(1993).
 RN [3]
 RP SEQUENCE OF 1765-2153 FROM N.A.
 RX MEDLINE=97075110; PubMed=8917536;
 RA Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,
 RA Martin D.I.;
 RT "Inhibition of granulocytic differentiation by mNotch1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).
 RN [4]
 RP FUNCTION.
 RX MEDLINE=99396706; PubMed=10393120;
 RA Hamada Y., Kadokawa Y., Okabe M., Ikawa M., Coleman J.R.,
 RA Tsujimoto Y.;
 RT "Mutation in ankyrin repeats of the mouse Notch2 gene induces early
 RT embryonic lethality.";
 RL Development 126:3415-3424(1999).
 RN [5]
 RP DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.
 RX MEDLINE=95333893; PubMed=7609614;
 RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;
 RT "Differential expression of Notch1 and Notch2 in developing and adult
 RT mouse brain.";
 RL Brain Res. Mol. Brain Res. 29:263-272(1995).
 RN [6]
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
 RX MEDLINE=21523956; PubMed=11518718;


```

FT REPEAT 1501 1533 LIN/NOTCH 2.
FT REPEAT 1825 1869 ANK 1.

Query Match 9.9%; Score 150.5; DB 1; Length 2470;
Best Local Similarity 24.3%; Pred. No. 0.0031;
Matches 53; Conservative 23; Mismatches 53; Indels 89; Gaps 12;

QY 31 ORMHENVSOGCGGATCSYNGCLSCPKPLFPALERIGMKQIGVCLSSCPGSGYCTRY 90
Db 945 QTMNECLSPCKNG-GTCSYVNSYTC-----TCPAGFHGVHC 982

QY 91 P-DINKTKCKADCDTCFNKNFC-----TKCKSGPYLHLGK 125
Db 983 ENNIDECR-----SSCFNGGTVDGINSFSLCPVGTGPPCLHDINESSNPLNAGT 1037

QY 126 CLDN-----CPEGLAANNHTMECVSIHV-CEVSENNPNSPCKTKGTCGPKRGGTTRV 177
Db 1038 CVDGLGYTCICPLGYTGKN---CQTLNVLCS-----RSPCKNGK-TC----- 1076

QY 178 REIIQHPKAGNLCPTNETKCTVQRKGC-----QKG 210
Db 1077 ---VQKARPHCLCPGPGWGCAYCDVLNVSCKAAALQKG 1111

RESULT 12
IGIR_RAT
ID IGIR_RAT STANDARD; PRT: 1370 AA.
AC P24062;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Insulin-like growth factor I receptor precursor (BC 2.7.1.112).
GN IGIRL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95277910; PubMed=7758167;
RA Du J., Delafontaine P.;
RT "Inhibition of vascular smooth muscle cell growth through antisense
transcription of a rat insulin-like growth factor I receptor cDNA.";
RL Circ. Res. 76:963-972(1995).
[2]
RN SEQUENCE OF 1-364 FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=50017496; PubMed=2477843;
RA Warner H., Woloschak M., Adamo M., Shen-Orr Z., Roberts C.T. Jr.,
RA Lerioth D.;
RT "Developmental regulation of the rat insulin-like growth factor I
receptor gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:7451-7455(1989).
[3]
RN SEQUENCE OF 913-1017 FROM N.A.
RX MEDLINE=92412145; PubMed=1530648;
RA Kurachi H., Jojo K., Ohta M., Kawasaki T., Itoh N.;
RT "A new member of the insulin receptor family, insulin
receptor-related receptor, is expressed preferentially in the
kidney.";
RL Biochem. Biophys. Res. Commun. 187:934-939(1992).
CC -1- FUNCTION: THIS RECEPTOR BINDS INSULIN-LIKE GROWTH FACTOR I (IGF I)
WITH A HIGH AFFINITY AND IGF II WITH A LOWER AFFINITY. IT HAS A
TYROSINE-PROTEIN KINASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC -1- SUBUNIT: Tetramer of 2 alpha and 2 beta chains linked by disulfide
bonds. The alpha chains contribute to the formation of the ligand-
binding domain, while the beta chain carries the kinase domain.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Belongs to the Tyr family of protein kinases. Insulin
receptor subfamily.

```

```

CC -1- SIMILARITY: Contains 2 fibronectin type III domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
or send an email to license@ebi.ac.uk).
CC
CC
DR EMBL; L29232; AAA41392.1; -
DR EMBL; M27293; AAA41384.1; -
DR HSP; P06213; IIRK.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR006211; Furin-like.
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow fac recep.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002011; RecepttyrkinasII.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR008266; Tyr_kinase_AS.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recept_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Prot_kinase; 1.
DR SMART; SM00060; FN3; 3.
DR SMART; SM00261; FU; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; Receptor; Transmembrane;
KW Glycoprotein; ATP-binding; Phosphorylation; Repeat; Signal.
FT SIGNAL 1 30
FT CHAIN 31 741
FT FT 742 1370
FT CHAIN 742 1370
FT DOMAIN 742 936
FT TRANSHEM 937 960
FT DOMAIN 961 1370
FT DOMAIN 608 829
FT DOMAIN 830 929
FT DOMAIN 1000 1275
FT NP_BIND 1006 1014
FT BINDING 1034 1034
FT ACT_SITE 1136 1136
FT DISULFID 215 224
FT DISULFID 219 230
FT DISULFID 231 239
FT DISULFID 235 248
FT DISULFID 251 260
FT DISULFID 264 276
FT DISULFID 282 303
FT DISULFID 307 321
FT DISULFID 324 328
FT CARBOHYD 51 51
FT CARBOHYD 102 102
FT CARBOHYD 135 135
FT CARBOHYD 145 145
FT CARBOHYD 245 245
FT CARBOHYD 314 314
FT CARBOHYD 418 418
FT CARBOHYD 439 439
FT CARBOHYD 535 535
FT CARBOHYD 608 608
FT CARBOHYD 623 623
FT CARBOHYD 641 641
FT CARBOHYD 748 748

```

```

-1- SIMILARITY: Contains 2 fibronectin type III domains.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
or send an email to license@ebi.ac.uk).
EMBL; L29232; AAA41392.1; -
EMBL; M27293; AAA41384.1; -
HSP; P06213; IIRK.
InterPro; IPR000494; EGFR_L_domain.
InterPro; IPR008957; FN_III-like.
InterPro; IPR003961; FN_III.
InterPro; IPR006211; Furin-like.
InterPro; IPR006212; Furin repeat.
InterPro; IPR009030; Grow fac recep.
InterPro; IPR000719; Prot_kinase.
InterPro; IPR002011; RecepttyrkinasII.
InterPro; IPR001245; Tyr_kinase.
InterPro; IPR008266; Tyr_kinase_AS.
Pfam; PF00041; fn3; 2.
Pfam; PF00757; Furin-like; 1.
Pfam; PF00069; pkinase; 1.
Pfam; PF01030; Recept_L_domain; 2.
PRINTS; PR00109; TYRKINASE.
PRODOM; PD000001; Prot_kinase; 1.
SMART; SM00060; FN3; 3.
SMART; SM00261; FU; 1.
SMART; SM00219; Tyrc; 1.
PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
Transferase; Tyrosine-protein kinase; Receptor; Transmembrane;
Glycoprotein; ATP-binding; Phosphorylation; Repeat; Signal.
SIGNAL 1 30
CHAIN 31 741
FT 742 1370
CHAIN 742 1370
DOMAIN 742 936
TRANSHEM 937 960
DOMAIN 961 1370
DOMAIN 608 829
DOMAIN 830 929
DOMAIN 1000 1275
NP_BIND 1006 1014
BINDING 1034 1034
ACT_SITE 1136 1136
DISULFID 215 224
DISULFID 219 230
DISULFID 231 239
DISULFID 235 248
DISULFID 251 260
DISULFID 264 276
DISULFID 282 303
DISULFID 307 321
DISULFID 324 328
CARBOHYD 51 51
CARBOHYD 102 102
CARBOHYD 135 135
CARBOHYD 145 145
CARBOHYD 245 245
CARBOHYD 314 314
CARBOHYD 418 418
CARBOHYD 439 439
CARBOHYD 535 535
CARBOHYD 608 608
CARBOHYD 623 623
CARBOHYD 641 641
CARBOHYD 748 748

```

```

FT CARBOHYD 757 757 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 765 765 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 901 901 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 914 914 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT MOD RES 1166 1166 PHOSPHORYLATION (AUTO-) (BY SIMILARITY) .
FT CONFLICT 985 986 AD -> PY (IN REF. 3) .
SQ SEQUENCE 1370 AA; 155395 MW; A5946897A41CB145 CRC64;

Query Match
Best Local Similarity 26.7%; Pred. No. 0.0023;
Matches 47; Conservative 17; Mismatches 46; Indels 66; Gaps 11;

QY 35 HPNVSGCGGCGATCSDYNGCLSCCKPRLPALRRIGMKQIGVCLSSCPGYY---GTRYP 91
Db 232 HPE---CLGSCHTPDNTTCAVRHYV-----KGVCVPACPPGTYRFGWR-- 275
QY 92 DINKTKKACDCTCNKPNFKTC-----KSGFYHLHGKCLDNCPGLEANN-HYMEC 143
Db 276 -----CVDKDFCANIPNAESSDSGDFVHDGECQECPSGFIRNSTQSMYC 321
QY 144 VSIHVCEVSEWNPSPCTKGGTKCGFKRGCTETRVREIIQHPSA-----KGNL 190
Db 322 IP---CE-----GPCP---RVCG---DEKKTITDSVTSQMLQGCTILKGNL 361

RESULT 13
NOTC_XENLA
ID NOTC_XENLA STANDARD; PRT; 2524 AA.
AC P21783;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch protein homolog precursor (XOTCH protein) .
GN XOTCH.
OS Xenopus laevis (African clawed frog) .
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=90385285; PubMed=2402639;
RA Coffman C., Harris W., Kintner C.;
RT "Xotch, the Xenopus homolog of Drosophila notch.",
RL Science 249:1438-1441(1990).
RN [2]
RP REVISIONS TO 1759-1782.
RA Kintner C.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DEVELOPMENTAL STAGE: Expressed almost uniformly in early embryos.
CC -1- SIMILARITY: Belongs to the NOTCH family.
CC -1- SIMILARITY: Contains 36 EGF-like domains.
CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
CC -1- SIMILARITY: Contains 6 ANK repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; M33874; AAB02039.1; -
CC HSSP; P00740; 1EDM.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx hydroxyl_s.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR002049; Laminin_EGF.

```

```

DR InterPro; IPR008297; Notch.
DR InterPro; IPR008000; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 36.
DR Pfam; PF00066; notch; 3.
DR PIRSF; PIRSF002279; Notch; 1.
DR PRINTS; PR00010; EGF_BLOOD.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00179; EGF_CA; 24.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS02937; ANK_REPEAT; 1.
DR PROSITE; PS00088; ANK_REPEAT; 4.
DR PROSITE; PS00010; ASX_HYDROXYL; 23.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 29.
DR PROSITE; PS00026; EGF_3; 36.
DR PROSITE; PS01187; EGF_CA; 21.
KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
PT SIGNAL 1 19 POTENTIAL.
PT CHAIN 20 2524 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG.
PT TRANSMEM 1729 1750 EXTRACELLULAR (POTENTIAL) .
PT DOMAIN 1751 2524 CYTOPLASMIC (POTENTIAL) .
PT DOMAIN 20 57 EGF-LIKE 1.
PT DOMAIN 58 99 EGF-LIKE 2.
PT DOMAIN 102 140 EGF-LIKE 3.
PT DOMAIN 141 177 EGF-LIKE 4.
PT DOMAIN 179 215 EGF-LIKE 5.
PT DOMAIN 217 254 EGF-LIKE 6.
PT DOMAIN 256 292 EGF-LIKE 7.
PT DOMAIN 294 332 EGF-LIKE 8.
PT DOMAIN 334 370 EGF-LIKE 9.
PT DOMAIN 371 409 EGF-LIKE 10.
PT DOMAIN 411 449 EGF-LIKE 11.
PT DOMAIN 451 487 EGF-LIKE 12.
PT DOMAIN 489 525 EGF-LIKE 13.
PT DOMAIN 527 563 EGF-LIKE 14.
PT DOMAIN 565 600 EGF-LIKE 15.
PT DOMAIN 602 638 EGF-LIKE 16.
PT DOMAIN 640 675 EGF-LIKE 17.
PT DOMAIN 677 713 EGF-LIKE 18.
PT DOMAIN 715 750 EGF-LIKE 19.
PT DOMAIN 752 788 EGF-LIKE 20.
PT DOMAIN 790 826 EGF-LIKE 21.
PT DOMAIN 828 866 EGF-LIKE 22.
PT DOMAIN 868 904 EGF-LIKE 23.
PT DOMAIN 906 942 EGF-LIKE 24.
PT DOMAIN 944 980 EGF-LIKE 25.
PT DOMAIN 982 1018 EGF-LIKE 26.
PT DOMAIN 1020 1056 EGF-LIKE 27.
PT DOMAIN 1058 1094 EGF-LIKE 28.
PT DOMAIN 1096 1142 EGF-LIKE 29.
PT DOMAIN 1144 1180 EGF-LIKE 30.
PT DOMAIN 1182 1218 EGF-LIKE 31.
PT DOMAIN 1220 1264 EGF-LIKE 32.
PT DOMAIN 1266 1304 EGF-LIKE 33.
PT DOMAIN 1306 1346 EGF-LIKE 34.
PT DOMAIN 1347 1383 EGF-LIKE 35.
PT DOMAIN 1386 1424 EGF-LIKE 36.
PT REPEAT 1441 1478 LIN/NOTCH 1.
PT REPEAT 1479 1520 LIN/NOTCH 2.
PT REPEAT 1521 1560 LIN/NOTCH 3.
PT REPEAT 1562 1599 ANK 1.
PT REPEAT 1600 1637 ANK 2.
PT REPEAT 1638 1675 ANK 3.
PT REPEAT 1676 1713 ANK 4.
PT REPEAT 1714 1751 ANK 5.
PT REPEAT 1752 1789 ANK 6.
PT DISULFID 22 35 BY SIMILARITY.
PT DISULFID 29 45 BY SIMILARITY.

```

```
FT DISULFID 47 56 BY SIMILARITY.
FT DISULFID 62 74 BY SIMILARITY.
FT DISULFID 68 87 BY SIMILARITY.
FT DISULFID 89 98 BY SIMILARITY.
FT DISULFID 106 117 BY SIMILARITY.
FT DISULFID 111 128 BY SIMILARITY.
FT DISULFID 130 139 BY SIMILARITY.
FT DISULFID 145 156 BY SIMILARITY.
FT DISULFID 150 165 BY SIMILARITY.
FT DISULFID 167 176 BY SIMILARITY.
FT DISULFID 183 194 BY SIMILARITY.
FT DISULFID 188 203 BY SIMILARITY.
FT DISULFID 205 214 BY SIMILARITY.
FT DISULFID 221 232 BY SIMILARITY.
FT DISULFID 226 242 BY SIMILARITY.
FT DISULFID 244 253 BY SIMILARITY.
FT DISULFID 260 271 BY SIMILARITY.
FT DISULFID 265 280 BY SIMILARITY.
FT DISULFID 282 291 BY SIMILARITY.
FT DISULFID 298 311 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 332 331 BY SIMILARITY.
FT DISULFID 338 349 BY SIMILARITY.
FT DISULFID 343 358 BY SIMILARITY.
FT DISULFID 360 369 BY SIMILARITY.
FT DISULFID 375 386 BY SIMILARITY.
FT DISULFID 380 397 BY SIMILARITY.
FT DISULFID 399 408 BY SIMILARITY.
FT DISULFID 415 428 BY SIMILARITY.
FT DISULFID 422 437 BY SIMILARITY.
FT DISULFID 439 448 BY SIMILARITY.
FT DISULFID 455 466 BY SIMILARITY.
FT DISULFID 460 475 BY SIMILARITY.
FT DISULFID 477 486 BY SIMILARITY.
FT DISULFID 493 504 BY SIMILARITY.
FT DISULFID 498 513 BY SIMILARITY.
FT DISULFID 515 524 BY SIMILARITY.
FT DISULFID 531 542 BY SIMILARITY.
FT DISULFID 536 551 BY SIMILARITY.
FT DISULFID 553 562 BY SIMILARITY.
FT DISULFID 569 579 BY SIMILARITY.
FT DISULFID 574 588 BY SIMILARITY.
FT DISULFID 590 599 BY SIMILARITY.
FT DISULFID 606 617 BY SIMILARITY.
FT DISULFID 611 626 BY SIMILARITY.
FT DISULFID 628 637 BY SIMILARITY.
FT DISULFID 644 654 BY SIMILARITY.
FT DISULFID 649 663 BY SIMILARITY.
FT DISULFID 665 674 BY SIMILARITY.
FT DISULFID 681 692 BY SIMILARITY.
FT DISULFID 686 701 BY SIMILARITY.
FT DISULFID 703 712 BY SIMILARITY.
FT DISULFID 719 729 BY SIMILARITY.
FT DISULFID 724 738 BY SIMILARITY.
FT DISULFID 740 749 BY SIMILARITY.
FT DISULFID 756 767 BY SIMILARITY.
FT DISULFID 761 776 BY SIMILARITY.
FT DISULFID 778 787 BY SIMILARITY.
FT DISULFID 794 805 BY SIMILARITY.
FT DISULFID 799 814 BY SIMILARITY.
FT DISULFID 816 825 BY SIMILARITY.
FT DISULFID 832 843 BY SIMILARITY.
FT DISULFID 837 854 BY SIMILARITY.
FT DISULFID 856 865 BY SIMILARITY.
FT DISULFID 872 883 BY SIMILARITY.
FT DISULFID 877 892 BY SIMILARITY.
FT DISULFID 894 903 BY SIMILARITY.
FT DISULFID 910 921 BY SIMILARITY.
FT DISULFID 915 930 BY SIMILARITY.
FT DISULFID 932 941 BY SIMILARITY.
FT DISULFID 986 997 BY SIMILARITY.
FT DISULFID 991 1006 BY SIMILARITY.
FT DISULFID 1008 1017 BY SIMILARITY.
```

```
FT DISULFID 1024 1035 BY SIMILARITY.
FT DISULFID 1029 1044 BY SIMILARITY.
FT DISULFID 1046 1055 BY SIMILARITY.
FT DISULFID 1062 1073 BY SIMILARITY.
FT DISULFID 1067 1082 BY SIMILARITY.
FT DISULFID 1084 1093 BY SIMILARITY.
FT DISULFID 1100 1121 BY SIMILARITY.
FT DISULFID 1115 1130 BY SIMILARITY.

Query Match 9.8%; Score 148.5; DB 1; Length 2524;
Best Local Similarity 22.5%; Pred. No. 0.0044;
Matches 73; Conservative 32; Mismatches 102; Indels 117; Gaps 19;

Qy 34 MHPNVSQCGGGATCSYNGCLSKCPRLPALERIGM-----KQIGV 76
Db 869 MNECVNPRNG-ATQNTNGSYKCNCKPGYTGNCENMDIDCQPNPCHNGSGSDGINM 927
Qy 77 CLSCLSPGYYGTR-YPDINKCTK--CK--ADCDTCFNKNFCTKCKGFF-LHL----- 123
Db 928 PFCNCPAGFRGPKCEEDINECASNPCKNGANCTCVNSYTC-CQPGSGIHCSNTPDC 986
Qy 124 -----GKCLD-----NCPGLEAN--NHTM-ECVSI----- 146
Db 987 TESSCFNGGTCIDGINTFTCCPPGFTGSGYCOHDINECDKPKCLNGGTCDPSYGYKCTC 1046
Qy 147 -----VHCE-VSEWNPWSPCTKKGK-----TCGFKG-----TETVRE 179
Db 1047 PQGYTGLANQNLVRWCSDSPCKGCKWQNNFYRCBCKSGMTGVYCDVPVSCVAAKQ 1106
Qy 180 ---LIQHSAGKGNLCPPNETRKTQVQKQKGERGKGRK-KPNK----- 226
Db 1107 QGVYDIVHLCRNSGMCVDTGTHFC-----RCQAGYTSYCEQVDECSNFCQNGATCTD 1161
Qy 227 ---GESKEAIPDSKSLSSKEIPE 247
Db 1162 YLGYSCECVAGYGVNCSBEINE 1185
```

RESULT 14

```
NTC2 RAT ID NTC2 RAT STANDARD; PRT; 2471 AA.
AC Q9QW30;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 2 precursor (Notch 2).
GN NOTCH2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
OX [1]
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93202015; Pubmed=1295745;
RA Weinmaster G., Roberts V.J., Lemke G.;
RT "Notch2: a second mammalian Notch gene.";
RL Development 116:931-941(1992).
RN [2]
RP TISSUE SPECIFICITY.
RX MEDLINE=21311789; Pubmed=11438922;
RA Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
functional roles for the Notch-DSL signaling system during brain
development.";
RL J. Comp. Neurol. 436:167-181(2001).
CC -I- FUNCTION: Functions as a receptor for membrane-bound ligands
Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs. May play an essential role in postimplantation
```

development, probably in some aspect of cell specification and/or differentiation (By similarity).

-!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-terminal fragment N(SC) which are probably linked by disulfide bonds (By similarity).

-!- SUBCELLULAR LOCATION: Type I membrane protein. Following proteolytical processing NICD is translocated to the nucleus.

-!- TISSUE SPECIFICITY: Highly expressed in the spleen and choroid plexus in the brain. Expressed in postnatal central nervous system (CNS) germinal zones and, in early postnatal life, within numerous cells throughout the CNS. It is more highly localized to ventricular germinal zones. Also found in the heart, liver and kidney.

-!- DEVELOPMENTAL STAGE: Expressed in the brain during E14 and E17.

-!- PTM: Synthesized in the endoplasmic reticulum as an inactive form which is proteolytically cleaved by a furin-like convertase in the trans-Golgi network before it reaches the plasma membrane to yield an active, ligand-accessible form. Cleavage results in a C-terminal fragment N(TM) and a N-terminal fragment N(SC). Following ligand binding, it is cleaved by TNP-alpha converting enzyme (TACE) to yield a membrane-associated intermediate fragment called notch extracellular truncation (NEXT). This fragment is then cleaved by presenilin dependent gamma-secretase to release a notch-derived peptide containing the intracellular domain (NICD) from the membrane (By similarity).

-!- PTM: Phosphorylated (By similarity).

-!- SIMILARITY: Belongs to the NOTCH family.

-!- SIMILARITY: Contains 35 EGF-like domains.

-!- SIMILARITY: Contains 2 lin/Notch repeats.

-!- SIMILARITY: Contains 6 ANK repeats.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M93661; AAK13558.1; --

PIR; A49128; A49128.

HSSP; P00743; 1CCF.

InterPro: IPR002110; ANK.

InterPro: IPR000152; Asx_hydroxyl_S.

InterPro: IPR000742; EGF_2

InterPro: IPR001881; EGF_Ca.

InterPro: IPR001438; EGF_II.

InterPro: IPR006209; EGF_like.

InterPro: IPR002049; Laminin_EGF.

InterPro: IPR008297; Notch.

InterPro: IPR008000; Notch_dom.

Pfam; PF00023; ank; 6.

Pfam; PF00008; EGF; 35.

Pfam; PF00066; notch; 2.

Pfam; PF002279; Notch; 1.

PRINTS; PR00010; EGF_BLOOD.

PRINTS; PR00011; EGF_LAMININ.

PRINTS; PR01452; NOTCH.

SMART; SM00248; ANK; 6.

SMART; SM00179; EGF_Ca; 24.

SMART; SM00004; NL; 2.

PROSITE; PS50297; ANK_REPEAT_REGION; 1.

PROSITE; PS50088; ANK_REPEAT; 4.

PROSITE; PS00010; ASX_HYDROXYL; 22.

PROSITE; PS00022; EGF_1; 34.

PROSITE; PS01186; EGF_2; 26.

PROSITE; PS50026; EGF_3; 35.

PROSITE; PS01187; EGF_Ca; 22.

Receptor; Transcription regulation; Activator; Differentiation;

Developmental protein; Repeat; ANK repeat; EGF-like domain;

Transmembrane; Glycoprotein; Signal; Phosphorylation.

SIGNAL 1 25 POTENTIAL.

CHAIN 26 2471 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.

CHAIN	FT	1666	2471	NOTCH EXTRACELLULAR TRUNCATION (BY SIMILARITY).
CHAIN	FT	1697	2471	NOTCH INTRACELLULAR DOMAIN (BY SIMILARITY)
DOMAIN	FT	26	1677	EXTRACELLULAR (POTENTIAL).
TRANSMEM	FT	1678	1698	POTENTIAL.
DOMAIN	FT	1699	2471	CYTOPLASMIC (POTENTIAL).
DOMAIN	FT	26	63	EGF-LIKE 1.
DOMAIN	FT	64	102	EGF-LIKE 2.
DOMAIN	FT	105	143	EGF-LIKE 3.
DOMAIN	FT	144	180	EGF-LIKE 4.
DOMAIN	FT	182	219	EGF-LIKE 5.
DOMAIN	FT	221	258	EGF-LIKE 6.
DOMAIN	FT	260	296	EGF-LIKE 7.
DOMAIN	FT	298	336	EGF-LIKE 8.
DOMAIN	FT	338	374	EGF-LIKE 9.
DOMAIN	FT	375	413	EGF-LIKE 10.
DOMAIN	FT	415	454	EGF-LIKE 11.
DOMAIN	FT	456	492	EGF-LIKE 12.
DOMAIN	FT	494	530	EGF-LIKE 13.
DOMAIN	FT	532	568	EGF-LIKE 14.
DOMAIN	FT	570	605	EGF-LIKE 15.
DOMAIN	FT	607	643	EGF-LIKE 16.
DOMAIN	FT	645	680	EGF-LIKE 17.
DOMAIN	FT	682	718	EGF-LIKE 18.
DOMAIN	FT	720	755	EGF-LIKE 19.
DOMAIN	FT	757	793	EGF-LIKE 20.
DOMAIN	FT	795	831	EGF-LIKE 21.
DOMAIN	FT	833	871	EGF-LIKE 22.
DOMAIN	FT	873	909	EGF-LIKE 23.
DOMAIN	FT	911	947	EGF-LIKE 24.
DOMAIN	FT	949	985	EGF-LIKE 25.
DOMAIN	FT	987	1023	EGF-LIKE 26.
DOMAIN	FT	1025	1061	EGF-LIKE 27.
DOMAIN	FT	1063	1099	EGF-LIKE 28.
DOMAIN	FT	1101	1147	EGF-LIKE 29.
DOMAIN	FT	1149	1185	EGF-LIKE 30.
DOMAIN	FT	1187	1223	EGF-LIKE 31.
DOMAIN	FT	1225	1262	EGF-LIKE 32.
DOMAIN	FT	1264	1302	EGF-LIKE 33.
DOMAIN	FT	1304	1343	EGF-LIKE 34.
DOMAIN	FT	1374	1412	EGF-LIKE 35.
DOMAIN	FT	1645	1648	POLY-ALA.
DOMAIN	FT	1994	1997	POLY-SER.
DOMAIN	FT	2426	2429	POLY-GLY.
REPEAT	FT	2446	2451	LIN/NOTCH 1.
REPEAT	FT	1420	1456	LIN/NOTCH 2.
REPEAT	FT	1503	1535	ANK 1.
REPEAT	FT	1827	1871	ANK 2.
REPEAT	FT	1876	1905	ANK 3.
REPEAT	FT	1909	1939	ANK 4.
REPEAT	FT	1943	1972	ANK 5.
REPEAT	FT	1976	2005	ANK 6.
REPEAT	FT	2009	2036	BY SIMILARITY.
DISULFID	FT	28	41	BY SIMILARITY.
DISULFID	FT	35	51	BY SIMILARITY.
DISULFID	FT	53	62	BY SIMILARITY.
DISULFID	FT	68	79	BY SIMILARITY.
DISULFID	FT	73	90	BY SIMILARITY.
DISULFID	FT	92	101	BY SIMILARITY.
DISULFID	FT	109	121	BY SIMILARITY.
DISULFID	FT	115	131	BY SIMILARITY.
DISULFID	FT	133	142	BY SIMILARITY.
DISULFID	FT	148	159	BY SIMILARITY.
DISULFID	FT	153	168	BY SIMILARITY.
DISULFID	FT	170	179	BY SIMILARITY.
DISULFID	FT	186	198	BY SIMILARITY.
DISULFID	FT	192	207	BY SIMILARITY.
DISULFID	FT	209	218	BY SIMILARITY.
DISULFID	FT	225	236	BY SIMILARITY.
DISULFID	FT	230	246	BY SIMILARITY.
DISULFID	FT	248	257	BY SIMILARITY.
DISULFID	FT	264	275	BY SIMILARITY.

Query Match 9.7%; Score 147.5; DB 1; Length 2471;
 Best Local Similarity 24.8%; Pred. No. 0.005;
 Matches 54; Conservative 23; Mismatches 52; Indels 89; Gaps 13;

QY 31 QRMHNVSGCGGATCSDYNGCTSCPKRLPFALERIGKQIGVCLLSCPSGYGTRY 90
 DB 947 QTDNECLSPCKNG-GTCSDYVNSYTC-----TCPAGFHVHC 984
 QY 91 P-DINKCTCKADCDCNENKFC-----TKCKSGFYHLGK 125
 DB 985 ENNIDECTE-----SSCFNGGTCVDGINSFSLCPVGFTGPPFLHDINECSSNPCLNSGT 1039
 QY 126 CLD-----NCEGLEANNHTMECVSIHV-CEVSENNPWSPTCKGKTCGFKRGTRV 177
 DB 1040 CVDGLGYRCTCPLGYTGKN-----CQTLVNLG-----SP-SPCKNG-TCA----- 1079
 QY 178 REIIQHPKAGKNCPTNETRKTCTVQRKCC-----QNG 210
 DB 1080 ----QEKARPRCLCPGWDGAYCDVLNVSKAAALQNG 1113

RESULT 15

BL14_CABEL
 ID BL14_CABEL STANDARD; PRT; 943 AA.
 AC P51559; O44762; O44763; O44764; O44765; O44766;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Endoprotease bli-4 precursor (EC 3.4.21.-) (Blisterase) (Blistered cuticle protein 4).
 DE BLI-4 OR KPC-4 OR K04F10.4.
 GN Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN [1]
 RN SEQUENCE FROM N.A. (ISOFORMS A; B; C AND D), AND FUNCTION.
 RC STRAIN=Bristol N2;
 RC MEDLINE=95293228; PubMed=7774813;
 RA Thacker C., Peters K.W., Srayko M., Rose A.M.;
 RT distinct kex2/subtilisin-like endoproteases essential for early development and adult morphology.;
 RT Genes Dev. 9:956-971(1995).
 RN [2]
 RN SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC STRAIN=Bristol N2;
 RA Latreille P., Wameile P.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN MUTAGENESIS.

RX MEDLINE=20363723; PubMed=10903434;
 RA Thacker C., Srayko M., Rose A.M.;
 RT "Mutational analysis of bli-4/kpc-4 reveals critical residues required for proprotein convertase function in C. elegans.";
 RL Gene 252:15-25(2000).
 CC -1- FUNCTION: The Kex2/subtilisin-like proteinase activity of this enzyme is required for the normal production of adult cuticle. There is functional redundancy between the isoforms. Vital for embryonic and larval development.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=5;
 CC Name=D; Synonyms=d;
 CC IsoId=P51559-1; Sequence=Displayed;
 CC Name=A; Synonyms=a;
 CC IsoId=P51559-2; Sequence=VSP_005416, VSP_005419;
 CC Name=B; Synonyms=b;
 CC IsoId=P51559-3; Sequence=VSP_005418, VSP_005421;
 CC Name=C; Synonyms=c;
 CC IsoId=P51559-4; Sequence=VSP_005422, VSP_005423;
 CC Name=E;
 CC IsoId=P51559-5; Sequence=VSP_005417, VSP_005420;
 CC Note=No experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: In larvae and adults, expressed in all hypodermal cells, vulva and ventral nerve cords.
 CC -1- DEVELOPMENTAL STAGE: Expression starts at embryo two-fold stage through to adults.
 CC -1- SIMILARITY: Belongs to peptidase family S8. Purin subfamily.
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to a frameshift in position 651.
 CC -1- CAUTION: Ref.2 (AAB96754 and AAB96757) sequence differs from that shown due to erroneous gene model prediction.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).
 CC -----
 CC EMBL; L29438; AAA98750.1; ALT_FRAME.
 CC EMBL; L29439; AAA98751.1; ALT_FRAME.
 CC EMBL; L29440; AAA98752.1; ALT_FRAME.
 CC EMBL; AF039719; AAB96753.1; -
 CC EMBL; AF039719; AAB96754.1; ALT_SEQ.
 CC EMBL; AF039719; AAB96755.1; -
 CC EMBL; AF039719; AAB96756.1; -
 CC EMBL; AF039719; AAB96757.1; ALT_SEQ.
 CC MEROPS; S08.031; -
 CC WormPep; K04F10.4a; CE11728.
 CC WormPep; K04F10.4b; CE11730.
 CC WormPep; K04F10.4c; CE11732.
 CC WormPep; K04F10.4d; CE11734.
 CC WormPep; K04F10.4e; CE11736.
 CC GO; 0016021; C: integral to membrane; NAS.
 CC GO; 0004352; P: serine-type endopeptidase activity; NAS.
 CC GO; 0007592; P: cuticle biosynthesis (sensu Invertebrata); IMP.
 CC InterPro; IPR006212; Furin repeat.
 CC InterPro; IPR009030; Grow fac recep.
 CC InterPro; IPR000209; Peptidase_S8.
 CC InterPro; IPR002884; Peptidase_S8B.
 CC InterPro; IPR009020; Protease_inhib.
 CC Pfam; PF01483; P: proprotein; I.
 CC Pfam; PF00082; Peptidase_S8; I.
 CC PRINTS; PR00723; SUBTILISIN.
 CC ProDom; PD000717; P domain; 1.
 CC SMART; SM00261; FU; 3.
 CC PROSITE; PS00136; SUBTILASE_ASP; 1.
 CC PROSITE; PS00137; SUBTILASE_HIS; 1.
 CC PROSITE; PS00138; SUBTILASE_SER; 1.
 CC Hydroxylase; Serine protease; Glycoprotein; Calcium-binding; Zymogen;
 KW Transmembrane; Signal; Alternative splicing.
 FT SIGNAL 1 20 POTENTIAL.

FT	PROPEP	21	116	BY SIMILARITY.
FT	CHAIN	117	943	ENDOPROTEASE BLI-4.
FT	ACT_SITE	202	202	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	241	241	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	415	415	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	TRANSNEM	417	437	POTENTIAL.
FT	DOMAIN	438	684	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	713	857	CYS-RICH.
FT	CARBOHYD	195	195	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARSPLIC	658	670	VEESARSPFDLT -> ILITIAIHLVVNA (in isoform A).
FT				/FTid=VSP_005416.
FT	VARSPLIC	658	699	VEESARSPFDLTSGWKLSCDCNCGGCTESSATSCPAYKH
FT				L -> LCFNPNESQPTCEVLCRCPGVINFOIEKLKST.
FT				/FTid=VSP_005417.
FT	VARSPLIC	658	730	VEESARSPFDLTSGWKLSCDCNCGGCTESSATSCPAYKH
FT				LTQTRNKGSGGPKVCVKQCDDTYYLDGDKCM -> GDEVV
FT				ERIRHNHEVTLIEBSHNWHEHARLQSLQELNSSRTHSPL
FT				YSPTFKQPIFLIILVCIFDAIHRQPAV (in isoform B).
FT				/FTid=VSP_005418.
FT	VARSPLIC	671	943	Missing (in isoform A).
FT				/FTid=VSP_005419.
FT	VARSPLIC	700	943	Missing (in isoform B).
FT				/FTid=VSP_005420.
FT	VARSPLIC	731	943	Missing (in isoform B).
FT				/FTid=VSP_005421.
FT	VARSPLIC	780	828	ESNLVQAKCIWRKDLCDGYYINAVGKDCDLCDSCTCTAP
FT				GPMSCEKC -> AENFDPCAKNSESGRDTYVVRFKPSAF
FT				KDYCNLKLVDLDPFISLLF (in isoform C).
FT				/FTid=VSP_005422.
FT	VARSPLIC	829	943	Missing (in isoform C).
FT				/FTid=VSP_005423.
FT	CONFLICT	153	153	A -> R (IN REF. 2).
FT	CONFLICT	346	346	S -> R (IN REF. 2).
FT	SEQUENCE	943 AA;	103146 MW;	401E009B6C46AD7E CRC64;
FT				
QY				Query Match
QY				Best Local Similarity
QY				Matches
QY				51; Conservative
QY				9.6%; Score 145; DB 1; Length 943;
QY				25.6%; Pred No. 0.0031;
QY				77; Indels 50; Gaps 117;
QY				
QY				32 RRMHPNVSQ-----COGSCATCSYNGCLSKPRLFFALERIGMKQIGVGLSSCPSG 84
QY				663 RSSFPDLTSGWKLSCDCNCGCTESSATSCPAYK-HLTQLRNKGGSGP-KCVQKCDT 720
QY				85 YGTRYPDINKCTKACDCTCFKNNPCTCKSGFYL-----HL-GKCLDNCPEGLE 135
QY				721 Y-----YLDGDKCMCSSCHCTCTKAEVCETCPGSLLLIDVNMPHYDHGKCVESCPFGLV 776
QY				136 ANNHTEMCVSIHVCEVSEWNPWSPTCKGKTGPKRGTTETVREIIOHPSAKMLCPPTN 195
QY				777 ADYES-----NLVQAK-----CIWRKDLG-----DGYVINAVGKCDLDCSSC 814
QY				196 ETRKCT-----VQRKKCKGK 210
QY				815 ET--CTAPGPMSCEKCSKG 831

Search completed: June 29, 2004, 17:00:16
Job time : 12.2828 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 16:55:24 ; Search time 34.7876 Seconds
(without alignments)
2476.067 Million cell updates/sec

Title: US-09-894-912A-13

Perfect score: 1516

Sequence: 1 MGLRLISLWFLINPMEYI.....QQKKRVQDKSVSVSTVH 273

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL 25:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rvirus:*
 - 16: sp_bacteriap:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1505	99.3	272	4	Q9BX4
2	1452	95.8	292	4	Q96K87
3	1153	76.1	217	11	Q9CSB2
4	1064	70.2	224	11	Q8BVW2
5	644	42.5	255	11	Q92132
6	626.5	41.3	236	4	Q8N7L5
7	576.5	38.0	243	11	Q8BFU0
8	472	31.1	224	4	Q9UGB2
9	459.5	30.3	176	4	Q8N6X6
10	417.5	27.5	152	11	Q7TPX3
11	283.5	18.7	138	11	Q8BJ73
12	192	12.7	129	5	Q26489
13	179.5	11.8	913	13	Q8AY18
14	177	11.7	1101	5	Q964D2
15	176.5	11.6	1376	5	Q8SZS2
16	176	11.6	1074	5	Q964D1

17	172	11.3	915	11	Q91VK0
18	167.5	11.0	296	11	O35171
19	167.5	11.0	826	11	O8CFZ2
20	167.5	11.0	932	11	O62030
21	163	10.8	3869	5	O86PQ3
22	162	10.7	913	4	O96EP4
23	161	10.6	440	5	O18003
24	159	10.5	503	5	O9U018
25	157.5	10.4	803	13	O42114
26	156	10.3	898	5	O76822
27	155	10.2	808	13	O42113
28	153.5	10.1	1362	13	O9PVZ4
29	151	10.0	238	5	O76510
30	149.5	9.9	4007	4	O86XX4
31	149	9.8	1371	11	O9QVW4
32	148	9.8	548	5	O9GQ45
33	148	9.8	4010	11	O80T14
34	147.5	9.7	802	13	O9W770
35	146.5	9.7	365	11	O924Y6
36	146.5	9.7	807	4	O9HCB6
37	145	9.6	724	4	O94862
38	145	9.6	807	4	O8NCD7
39	144.5	9.5	660	5	O23832
40	144.5	9.5	2189	5	O9BI05
41	144	9.5	807	6	O9GLX9
42	143.5	9.5	220	11	O99KR2
43	143.5	9.5	402	11	O8K2Q8
44	143.5	9.5	807	11	O8VCC9
45	143	9.4	435	5	O9GQ41

ALIGNMENTS

RESULT 1

Q9BX4 ID Q9BX4 PRELIMINARY; PRT; 272 AA.

AC Q9BX4; 01-JUN-2001 (T-EMBLrel. 17, Created)

DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (T-EMBLrel. 25, Last annotation update)

DE Thrombospondin.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Mao Y., Xie Y., Zhou Z., Zhao W., Wang S., Huang Y., Wang S.,

RA Tang R., Chen X., Wu C.,

RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUS=Placenta;

RA Strausberg R.;

RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.

RL ENBL; AP251057; AAK34947.1; -

DR ENBL; BC022367; AAH22367.1; -

DR Genew; HGNC:20866; THSD2.

DR InterPro; IPR006212; Furin_repeat.

DR InterPro; IPR009030; Grow_fac_recep.

DR InterPro; IPR000884; TSPI.

DR Pfam; PF00090; tspl_1; 1.

DR SMART; SM00261; FU; 2.

DR SMART; SM00209; TSPI; 1.

DR PROSITE; PSS0092; TSPI; 1.

SQ SEQUENCE 272 AA; 30928 MW; CACAEC6B7E781189 CRC64;

Query Match 99.3%; Score 1505; DB 4; Length 272;

Best Local Similarity 100.0%; Pred. NO. 4.3e-130; Indels 0;

Matches 271; Conservative 0; Mismatches 0; Gaps 0;

QY 3 HLRISLWFLINPMEYIQQKKRVQDKSVSVSTVH 62


```
Db 2 HLLRISWLFILFMFYIGSNASGRQRMRHNPVSGQGCGCATCSDYNGCLSCPKRL 61
Qy 63 PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKTKCKADCDTCFNNKFTCKKSGFYLH 122
Db 62 PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKTKCKADCDTCFNNKFTCKKSGFYLH 121
Qy 123 LGKCLDNCPEGLNANNHTECVSIHVCEVSEWNPSPCTKKGKTCGFRGTETRVREIIQ 182
Db 122 LGKCLDNCPEGLNANNHTECVSIHVCEVSEWNPSPCTKKGKTCGFRGTETRVREIIQ 181
Qy 183 HPSAKGNLCPPTNETRKTQVQRKKQGERGKGRERKRPKNKGSKEAIPDSKSLESS 242
Db 182 HPSAKGNLCPPTNETRKTQVQRKKQGERGKGRERKRPKNKGSKEAIPDSKSLESS 241
Qy 243 KEIPEORENKOQKQKRVQDKQSVSVTVH 273
Db 242 KEIPEORENKOQKQKRVQDKQSVSVTVH 272

RESULT 2
Q96K87 PRELIMINARY; PRT; 292 AA.
AC Q96K87
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ14440.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiwa S., Komai P., Hara R., Takeuchi K.,
RA Arita M., Nabekura Y., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakanatsu A., Nakamura Y., Nagahara K., Masuho Y., Oshima A.;
RA "NEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027346; BAB55051.1; -.
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp_1; 1.
DR SMART; SM00261; FU; 2.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS50092; TSP1; 1.
KM Hypothetical protein.
SQ SEQUENCE 292 AA; 33243 MW; 01E2774AC3DA6F8 CRC64;

Query Match 95.8%; Score 1452; DB 4; Length 292;
Best Local Similarity 99.2%; Pred. No. 3.3e-125;
Matches 262; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 HLLRISWLFILFMFYIGSNASGRQRMRHNPVSGQGCGCATCSDYNGCLSCPKRL 62
Db 2 HLLRISWLFILFMFYIGSNASGRQRMRHNPVSGQGCGCATCSDYNGCLSCPKRL 61
Qy 63 PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKTKCKADCDTCFNNKFTCKKSGFYLH 122
Db 62 PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKTKCKADCDTCFNNKFTCKKSGFYLH 121
Qy 123 LGKCLDNCPEGLNANNHTECVSIHVCEVSEWNPSPCTKKGKTCGFRGTETRVREIIQ 182
Db 122 LGKCLDNCPEGLNANNHTECVSIHVCEVSEWNPSPCTKKGKTCGFRGTETRVREIIQ 181
Qy 183 HPSAKGNLCPPTNETRKTQVQRKKQGERGKGRERKRPKNKGSKEAIPDSKSLESS 242
Db 182 HPSAKGNLCPPTNETRKTQVQRKKQGERGKGRERKRPKNKGSKEAIPDSKSLESS 241
```

```
Qy 243 KEIPEORENKOQKQKRVQDKQSVSVTVH 266
Db 242 KEIPEORENKOQKQKRVQDKQSVSVTVH 265

RESULT 3
Q9CSB2 PRELIMINARY; PRT; 217 AA.
AC Q9CSB2
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE 2810459H04RIK protein (Fragment).
GN THSD2 OR 2810459H04RIK.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Komno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavani T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stabli P., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyszynski-Borja A., Yoshida K., Haegawa Y., Kawai H., Kohseki S.,
RA Hayashizaki Y.;
RA "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK013366; BAB28811.1; -.
DR MGD; MGI:1920030; Thsd2.
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp_1; 1.
DR SMART; SM00261; FU; 2.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS50092; TSP1; 1.
FT NON TER 217
SQ SEQUENCE 217 AA; 24304 MW; 0DCF938B9FB3FB7 CRC64;

Query Match 76.1%; Score 1153; DB 11; Length 217;
Best Local Similarity 93.1%; Pred. No. 6.7e-96;
Matches 201; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Qy 3 HLLRISWLFILFMFYIGSNASGRQRMRHNPVSGQGCGCATCSDYNGCLSCPKRL 62
Db 2 HLLRISWLFILFMFYIGSNASGRQRMRHNPVSGQGCGCATCSDYNGCLSCPKRL 61
Qy 63 PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKTKCKADCDTCFNNKFTCKKSGFYLH 122
Db 62 PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKTKCKADCDTCFNNKFTCKKSGFYLH 121
Qy 123 LGKCLDNCPEGLNANNHTECVSIHVCEVSEWNPSPCTKKGKTCGFRGTETRVREIIQ 182
Db 122 LGKCLDNCPEGLNANNHTECVSIHVCEVSEWNPSPCTKKGKTCGFRGTETRVREIIQ 181
Qy 183 HPSAKGNLCPPTNETRKTQVQRKKQGERGKGRERKRPKNKGSKEAIPDSKSLESS 218
Db 182 HPSAKGNLCPPTNETRKTQVQRKKQGERGKGRERKRPKNKGSKEAIPDSKSLESS 217
```

RESULT 4
 ID QBWV2 PRELIMINARY; PRT; 224 AA.
 AC QBWV2;
 DT 01-MAR-2003 (TREMREL. 23, Created)
 DT 01-MAR-2003 (TREMREL. 23, Last sequence update)
 DT 01-OCT-2003 (TREMREL. 25, Last annotation update)
 DE Thrombospondin homolog.
 GN THSD2 OR 2810459H04RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Skin;
 RC MEDLINE=22354683; PubMed=12466851;
 RA THE FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573 (2002).
 DR EMBL; AK076308; PAC36296.1; -.
 DR MGD; MGI:1920030; Thsd2.
 DR InterPro; IPR006212; Furin repeat.
 DR InterPro; IPR009030; Grow_fac_recep.
 DR SMART; SM00261; FU; 2.
 SQ SEQUENCE 224 AA; 25398 MW; BC1B083497CFB83 CRC64;
 Query Match 70.2%; Score 1064; DB 11; Length 224;
 Best Local Similarity 84.6%; Pred. No. 1e-89;
 Matches 187; Conservative 9; Mismatches 25; Indels 0; Gaps 0;
 QY 3 HRLISWLFILNFMFYIGSQNASRGRQRMRHFNVSQGGCGCATCSYNGCLSKPRL 62
 DB 2 HRLISCFILNFMFYIGSQNASRGRQRMRHFNVSQGGCGCATCSYNGCLSKPRL 61
 QY 63 FFLERIGMKOIGVCLSSCPGYYGTRYPDINKCTCKADCTCFNKNFCTCKSGFYHL 122
 DB 62 FFLERIGMKOIGVCLSSCPGYYGTRYPDINKCTCKADCTCFNKNFCTCKSGFYHL 121
 QY 123 LGRCLDNCPEGLRNNHMTCEVSIHVCEVSEWNPSPCTTKGKTCGFGRTETRVREIIQ 182
 DB 122 LGRCLDSCPEGLRNNHMTCEVSIHVCEVSEWNPSPCTTKGKTCGFGRTETRVREIIQ 181
 QY 183 HPSAKGNLCPPTNETRKTCTVQKCKGKRGKRGK 223
 DB 182 HPSAKGNVPVPPNQDRKLYTKKEVPKGRARKRGKRTKK 222
 RESULT 5
 QY2132 PRELIMINARY; PRT; 265 AA.
 ID QY2132
 AC QY2132;
 DT 01-MAY-1999 (TREMREL. 10, Created)
 DT 01-MAY-1999 (TREMREL. 10, Last sequence update)
 DT 01-OCT-2003 (TREMREL. 25, Last annotation update)
 DE Thrombospondin type 1 domain.
 GN RSPONDIN OR R-SPONDIN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC kamata T., Katsube K., Michikawa M., Yamada M., Mizusawa H.;
 RA "R-spondin, a novel thombospondin type 1 domain gene, expressed in
 RT the dorsal neural tube."
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB016768; BAA75640.1; -.
 DR MGD; MGI:2183426; Rspndin.
 DR

InterPro; IPR006212; Furin repeat.
 InterPro; IPR009030; Grow_fac_recep.
 Pfam; PF00090; tsp_1; 1.
 SMART; SM00261; FU; 2.
 SMART; SM00209; TSP1; 1.
 PROSITE; PS50092; TSP1; 1.
 SQ SEQUENCE 265 AA; 29331 MW; FFB8964743F5963 CRC64;
 Query Match 42.5%; Score 644; DB 11; Length 265;
 Best Local Similarity 46.5%; Pred. No. 4.1e-51;
 Matches 119; Conservative 36; Mismatches 77; Indels 24; Gaps 6;
 QY 6 LLSWLFILNFMFYIGSQNASRGRQRMRHFNVSQGGCGCATCSYNGCLSKPRLPFA 65
 DB 11 VLSWTHIA-----VGSRGVIGKGRQRISABGSOACAKGCELCSEVNGCLKSPKLFIL 63
 QY 66 LERIGMKOIGVCLSSCPGYYGTRYPDINKCTCKAD-CDTCFNKNFCTCKSGFYHLG 124
 DB 64 LERNDRQVGVCLPSCPPGYFDARNPDMNKC1KCKIEHCEACFSHNFTCKQCALYHLKG 123
 QY 125 KCLDNCPEGLRNNHMTCEVSIHVCEVSEWNPSPCTTKGKTCGFGRTETRVREIIQHP 184
 DB 124 RCVPCPEGSTAANGTMEGCSPAQCEMSEWSWPGCKSKKLGCGFKGSEBETRVLHAP 183
 QY 185 SAKGNLCPPTNETRKTCTVQKCKGKRGK-----RERKRPKNKSGEKAIPTDSKLE 240
 DB 184 GGDHTTCSDTKETRKTCTVTPCPBQKRRKGGQRRRANRHPARKNSKB--PRNS-- 239
 QY 241 SSKETPEQRENKQOK 256
 DB 240 -----RRHKGQOQ 247
 RESULT 6
 QY2175 PRELIMINARY; PRT; 236 AA.
 ID QY2175;
 AC QY2175;
 DT 01-OCT-2002 (TREMREL. 22, Created)
 DT 01-OCT-2002 (TREMREL. 22, Last sequence update)
 DT 01-OCT-2003 (TREMREL. 25, Last annotation update)
 DE Hypothetical protein FLJ40906.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=uterus;
 RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
 RA Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
 RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
 RA Sugano S., Nagahori K., Masuho Y., Nagai K., Isegai T.,
 RT "NEDO human cDNA sequencing project."
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK098225; BAC05263.1; -.
 DR InterPro; IPR006212; Furin repeat.
 DR InterPro; IPR009030; Grow_fac_recep.
 DR Pfam; PF00090; tsp_1; 1.
 DR SMART; SM00261; FU; 2.
 DR SMART; SM00209; TSP1; 1.
 DR PROSITE; PS50092; TSP1; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 236 AA; 25942 MW; 8D03803127EC5678 CRC64;
 Query Match 41.3%; Score 626.5; DB 4; Length 236;
 Best Local Similarity 48.1%; Pred. No. 1.5e-49;
 Matches 111; Conservative 36; Mismatches 69; Indels 15; Gaps 3;

QY 33 RMHPNVSCQGGCATCSDYNGCLSCPKPLFPALERIGMKQIGVCLSSCPGYYGTRYD 92
DB 4 RVSAEGSQACAGCELCSEVNGCLCKSPKLFILLERNDIRQVGLCLPSPGPGYFDARNPD 63
QY 93 INKCTKAD-CDTCFNKPFCTCKSGFYHLGKCLDNCPEGLAANNHTMECVSIHVCEV 151
DB 64 MKKCIKLEHCEACFSHFCTCKGGLYHLKRCYCPCACPEGSSAANGTMECSPPACCEV 123
QY 152 SEWNPSPCTKGGKTCGFKGTGTRVRIIOHPSAKGNLCPPNTRKCTVQKCKQKE 211
DB 124 SEWSPWGPCSKQQLCGFRGSRTERVRLHAPVGDHAACTDKTRCTVRRVPCPEQ 183
QY 212 RGKKGHRERKPKNG-----ESKEAIPDSKSLSSKEIPEQRENKQOQKR 258
DB 184 KRRKGQGRRENARNRLARKEBAGAGSR-----RRKQOQOQOQO 224

RESULT 7

Q8BFU0 PRELIMINARY; PRT; 243 AA.
AC Q8BFU0; 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical thrombospondin type I repeat.
GN 2610028F08RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RQ STRAIN=C57BL/6J; TISSUE=Bye, and Hippocampus;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium.
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK049891; BAC33974.1; --
DR EMBL; AK087485; BAC39893.1; --
DR MGD; MGI:1922667; 2610028F08RIK.
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR InterPro; IPR00884; TSPI.
DR SMART; SM00261; FU; 2.
DR SMART; SM00209; TSPI; 1.
DR PROSITE; PS50092; TSPI; 1.
KW Hypothetical protein.
SQ SEQUENCE 243 AA; 28275 MW; ED76A08D61012ED7 CRC64;

Query Match 38.0%; Score 576.5; DB 11; Length 243;
Best Local Similarity 45.0%; Pred. No. 5.9e-45;
Matches 108; Conservative 39; Mismatches 78; Indels 15; Gaps 5;

QY 6 LISWLPILNMFYIGSONASRGRRQRMHPNVSCQGGCATCSDYNGCLSCPKPLFPA 65
DB 5 LPSFALLIILNMDYSQC-QGRWRNRKRA-SYVSNPICKGLSCSDKNGSCRCQKLPFP 62
QY 66 LERIGHKQIGVCLSSCPGYYGTRYDINKCTKAD-CDTCFNKPFCTCKSGFYHLG 124
DB 63 LRREGHRYQVECLHSPGYYGHRAPDMNRCARCIENCDSCFSKDFCTCKVGYLH 122
QY 125 KCLDNCPEGLAANNHTMECVSIHVCEWNPSPCTKGGKTCGFKGTGTRVRIIOHP 184
DB 123 RCFDECPDGPAPLDETMECVE--GCEVHSEWGTCSRNRRTCGFKWGLETRQIVKPP 180
QY 185 SAKGNLCPPNTRKCTVQKCKQKGRGKGRERKKNKESKEAIPDSKSLSSKE 244
DB 181 AKDTIPCTTAESRRRCQAMRHCPGGKRTPKAKERNKKRR-----KLITERAQE 230

RESULT 8

Q9UGB2

ID Q9UGB2 PRELIMINARY; PRT; 224 AA.
AC Q9UGB2; 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DJ824P16.3 (Novel protein similar to mouse thrombospondin type 1 domain protein R-spondin) (Fragment).
GN DJ824P16.3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Blakey S.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL050325; CAB65783.3; --
DR Genew; HGNC:16175; C20orf182.
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR InterPro; IPR00884; TSPI.
DR SMART; SM00261; FU; 2.
DR SMART; SM00209; TSPI; 1.
DR PROSITE; PS50092; TSPI; 1.
FT NON TER 224
SQ SEQUENCE 224 AA; 25042 MW; 97D26AD34CDBF812 CRC64;

Query Match 31.1%; Score 472; DB 4; Length 224;
Best Local Similarity 43.3%; Pred. No. 2.1e-35;
Matches 87; Conservative 36; Mismatches 70; Indels 8; Gaps 5;

QY 29 RRQRMRHPNVSCQGGCATCSDYNGCLSCPKPLFPALERIGMKQIGVCLSSCPGYYGT 88
DB 22 RRRKQVGTGLGGNCT-GCIICSEENGSTCQQLFLFIRREGINQYKGLHDCPPGFGI 80
QY 89 RYDPINKCTKCKADCDTCFNKPFCTCKSGFYHLGKCLDNCPEGLAANNHTMECVSIHV 148
DB 81 RQGFVNRCKCGATCSCFSQDFCIRKQFYLKGLCLTPCPGTLAHQNTRECQ--E 138
QY 149 CEVSEWNPSPCTKGGKTCGFKGTGTRVRIIOHPSAKGNLCPPNTRKCTVQKCKQ 208
DB 139 CHLPGWGSPTCHNGKTCGSAWGLSRVREAGRAGHEEAATCQVLSERKCPQR-PC- 196
QY 209 KGERG---KGRERKRRKKNK 226
DB 197 PGRSPGQKGRKRRPRKDR 217

RESULT 9

Q8N6X6 PRELIMINARY; PRT; 176 AA.
ID Q8N6X6
AC Q8N6X6; 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to putative.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Straubeberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC027938; AAH27938.1; --
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR InterPro; IPR00884; TSPI.
DR SMART; SM00261; FU; 1.
DR SMART; SM00209; TSPI; 1.
DR PROSITE; PS50092; TSPI; 1.

```
SQ SEQUENCE 176 AA; 20409 MW; 0F83CCE1B2F8CA85 CRC64;

Query Match 30.3%; Score 459.5; DB 4; Length 176;
Best Local Similarity 45.1%; Pred. No. 2.3e-34;
Matches 79; Conservative 30; Mismatches 53; Indels 13; Gaps 3;

QY 71 MKQIGVCLSSCPSSGYGTRYPDINKTKCAD-CDTCFKNKPFCTKCKSGPYLHGLKCLDN 129
DB 1 MRQYGECLHSCPSGYGHRAPDMNRCARCIENCDSCFSDKFCCKCKVGYLHRCFDE 60

QY 130 CPBGLANNHTECVISVCEVSEWNPSPCTKGGTKTGPRGTRVRIIHPSPAGN 189
DB 61 CPDGFAPLEBTEBCE--GCEVHMSWEGTCSRNRRTCGFKGLETRQIVKPKVQDTI 118

QY 190 LCPPTNETKCTVORCKQCKGKGRKPKKPKNGESKEAIPDPSKSLESKE 244
DB 119 PCPTIABRRRCNMTMRHCPGGRTPRAKEKRNKKKKR-----KLIFRAQE 163

RESULT 10
Q7TPX3
ID Q7TPX3 PRELIMINARY; PRT; 152 AA.
AC Q7TPX3;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DE 2610028F08Rik protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Egg;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Altschul S.F., Collins F.S., Wagner L., Shennen G.D.,
RA Bhat N.K., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Wozney K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalek U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Egg;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC052844; AAHS2844.1; -.
SQ SEQUENCE 152 AA; 17658 MW; B74713789B2853B4 CRC64;

Query Match 27.5%; Score 417.5; DB 11; Length 152;
Best Local Similarity 52.3%; Pred. No. 1.4e-30;
Matches 79; Conservative 20; Mismatches 47; Indels 5; Gaps 4;

QY 6 LISWPIILNFWYIGSQVARSRCRRRRMHPNVSCQCGCATCSYNGCLSKCPPLPFA 65
DB 5 LFSFALIIILNCDYISQCG--GNRRRRNKRA--SYVSNPICKGLSCSDNGSCRCQCKLPPF 62

QY 66 LRIQGMQKIGVCLSSCPSSGYGTRYPDINKTKCAD-CDTCFKNKPFCTKCKSGPYLHGL 124
```

```
DB 63 LRREGMROYCECLHSCPSGYGHRAPDMNRCARCIENCDSCFSDKFCCKCKVGYLHRC 122
QY 125 KCLDNCPEGLANNHTECVISVCEVSEWN 155
DB 123 RCFDRCPDGSPAPLDETMECVE--GCEVGHWS 151

RESULT 11
Q8BJ73
ID Q8BJ73 PRELIMINARY; PRT; 138 AA.
AC Q8BJ73;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE DJ824F16.3
GN A930029KL9RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Retina;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK020904; BAC25643.1; -.
DR MGD; MGI:1924467; A930029KL9RIK.
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow_Fac_recep.
DR SMART; SM00261; FU; 2.
SQ SEQUENCE 138 AA; 15172 MW; FEFD7D949279D5DD CRC64;

Query Match 18.7%; Score 283.5; DB 11; Length 138;
Best Local Similarity 42.9%; Pred. No. 2.5e-18;
Matches 45; Conservative 24; Mismatches 35; Indels 1; Gaps 1;

QY 29 RRORRHNPVSCQCGCATCSYNGCLSKCPPLPALERIGMKQIGVCLSSCPSSGYGT 88
DB 22 RPKDQAGTGLGNGCT--GCVCSENGCSTCQRLFLFIRREGIRQYKGVCHDCPLGFFGI 80

QY 89 RYPDINKTKCADCDTCFKNKPFCTKCKSGPYLHGLKCLDNCPG 133
DB 81 RQGEANRCKKOGATCSYNGCLSKCPPLPALERIGMKQIGVCLSSCPSSGYGT 125

RESULT 12
Q26489
ID Q26489 PRELIMINARY; PRT; 1299 AA.
AC Q26489;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Endoprotease furin.
GN Furin.
OS Spodoptera frugiperda (Fall armyworm).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;
OC Noctuidae; Amphipyrinae; Spodoptera.
OX NCBI_TaxID=7108;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RA Cieplik M., Klenk H.;
RT "Cloning and functional characterization of FURIN from Spodoptera
RT frugiperda (sf9) cells."
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9.
CC -1- SIMILARITY: BELONGS TO HOMO B/P DOMAIN.
DR EMBL; Z68888; CAA93116.1; -.
```



```
QY 44 GGCATCSD---YNGCL---SC---KPLRPFAL---RIGMKQIGVCLSSCPGSGYGTGR 89
Db 773 GTQSCSDLSKYPCCKTDCNVDSRTGFIYATECSDFGSRSPSYNCTTCTASNYT-PR 831
QY 90 YPDINKCTKCAKCDCTCFNRNFTK-----CKSGFYHLGLKCLDNCPEGLEA 136
Db 832 EGEKNGCAKCDKCATCSDKDTCLTCADPLKVGSKDCKTGYTMSNGBC-----KPC 884
QY 137 NHTMECVSIVHCEVSEWNPWSPCTKKG-KTC--GPKRGTTETRVREIIQHPSAKGNLCPP 193
Db 885 TNHCSECSAACTVCSDYTKVISGNGCNSVDGF-----YFDEINGTCIPC 932
QY 194 TNETRKTCTVORKKCKGKRGKGRKK 220
Db 933 TSPTCKCVGVKQCEQEETGCNSEKK 959
```

RESULT 15

```
Q8SZS2 PRELIMINARY; PRT; 1376 RA.
AC Q8SZS2
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE LD30182p.
GN FUR2 OR CG4235 OR CG18734.
OS Drosophila melanogaster (Fruit fly).
OC Rukaryota; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Ceiniker S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC -1- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
DR EMBL; AY070553; AAL48024.1; -.
DR FlyBase; FBgn0004598; Fur2.
DR GO; GO:0004276; F:kin activity; IDA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR006212; Furin_repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR002884; Peptidase_S8B.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF01483; P_protein; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR ProDom; PD000717; P_domain; 1.
DR SMART; SM0261; FU_10.
DR PROSITE; PS00190; CYTOCHROME_C; 3.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Protease; Serine_protease.
SQ SEQUENCE 1376 AA; 149716 MW; B670BA89A3A88FB CRC64;
```

```
Query Match 11.6%; Score 176.5; DB 5; Length 1376;
Best Local Similarity 28.0%; Pred. No. 1.7e-07;
Matches 60; Conservative 24; Mismatches 77; Indels 53; Gaps 12;
QY 3 HLRLLSWLFIILNMEYIGSNASRGRRQRMRHNPWSQGGCATCSDY-WGCLSCKPR 61
Db 734 HLHVID-LAVLCQCPDGYFENS-----RNRTCVP-----CEPNCASQDHPFYCTSCDH 783
QY 62 LPFALERIGMKQIGVCLSSCPGSGYGTGRYPDINKCTKACDCTCF--NKNPCTKCKSGF 119
```

```
Db 784 LVMHEHK-----CYSACPDLDTYET---EDNKCAPCHSTCATCGPTDQDCITCRSSR 832
QY 120 YHLGLKCLDNCPEGLEBANNHTMECVSIVHCEVSEWNPWSPCTKKGKTCGPKRGTTETRV-R 178
Db 833 YAWQNKCLISCPDGFYADKRLKCM-----PCQEGCKTC-----TSNGVCS 873
QY 179 EIIQHPSAKGNLCPPTNETRKTCTVORKK-CQKGE 211
Db 874 ECLQWNT-----LNKRDKCVSGSGGSESE 899
```

Search completed: June 29, 2004, 17:02:12
Job time : 37.7876 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: June 29, 2004, 16:54:09 ; Search time 50.4502 Seconds
(without alignments)
1562.545 Million cell updates/sec

Title: US-09-894-912A-32
Perfect score: 1535
Sequence: 1 MHLRLISCFILNFMFYIG.....QQRARRDKQKXSVSVSTVH 279

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 29Jan04:.*
1: geneseqp1980s:.*
2: geneseqp1990s:.*
3: geneseqp2000s:.*
4: geneseqp2001s:.*
5: geneseqp2002s:.*
6: geneseqp2003as:.*
7: geneseqp2003bs:.*
8: geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1535	100.0	279	AAE13167	AAE13167 Mouse ste
2	1535	100.0	279	ABO44431	ABO44431 Mouse ste
3	1315.5	85.7	272	AAH99220	AAH99220 Human chr
4	1315.5	85.7	272	AAH78328	AAH78328 Human pro
5	1315.5	85.7	272	AAE13168	AAE13168 Human ste
6	1315.5	85.7	272	AAE13150	AAE13150 Human ste
7	1315.5	85.7	272	ABR62112	ABR62112 Human clo
8	1315.5	85.7	272	ABO44432	ABO44432 Human ste
9	1315.5	85.7	272	ABO44413	ABO44413 Human ste
10	1310.5	85.4	273	AAE13151	AAE13151 Human ste
11	1310.5	85.4	273	ABO44414	ABO44414 Human ste
12	1293	84.2	292	AAW85607	AAW85607 Secreted
13	1293	84.2	292	AAE13170	AAE13170 Human SCR
14	1293	84.2	292	ABP61846	ABP61846 Human pol
15	1293	84.2	292	ABR62114	ABR62114 Human sec
16	1293	84.2	292	ABO44434	ABO44434 Human SCR
17	1285	83.7	265	AAE13163	AAE13163 Human sec
18	1285	83.7	265	ABO44427	ABO44427 Human sec
19	1273	82.9	292	AAH93875	AAH93875 Human pro
20	1222.5	79.6	251	AAE13153	AAE13153 Human nat
21	1222.5	79.6	251	ABO44417	ABO44417 Human ste
22	1100	71.7	239	ABU52396	ABU52396 Human GPC
23	1045	68.1	195	ABU52398	ABU52398 Human GPC
24	1041	67.8	195	ABU52397	ABU52397 Human GPC
25	759	49.4	160	ABB11374	ABB11374 Human sec

26	759	49.4	160	4	AAH79312	AAH79312 Human pro
27	759	49.4	160	4	AAE13149	AAE13149 Human ste
28	759	49.4	160	6	ABO44415	ABO44415 Human ste
29	648	42.2	263	6	ABR62108	ABR62108 Secreted
30	648	42.2	263	6	ABR62115	ABR62115 Secreted
31	648	42.2	263	6	ABR58489	ABR58489 Human sec
32	648	42.2	263	7	ABO7919	ABO7919 Novel pro
33	640.5	41.7	243	6	ABR62110	ABR62110 Secreted
34	632	41.2	265	6	ABR62113	ABR62113 Mouse thr
35	629	41.0	229	4	AAE13162	AAE13162 Mouse thr
36	629	41.0	229	6	ABO44426	ABO44426 Human sec
37	571.5	37.2	243	5	ABG76508	ABG76508 DNA encod
38	569.5	37.1	243	5	ABR62106	ABR62106 Secreted
39	569.5	37.1	243	6	ABR62106	ABR62106 Secreted
40	565.5	36.8	243	6	AAE13166	AAE13166 Mouse SCR
41	565	36.8	250	6	ABR62101	ABR62101 Secreted
42	536	34.9	222	6	ABR62107	ABR62107 Secreted
43	531.5	34.6	229	6	ABR62102	ABR62102 Secreted
44	450	29.3	234	7	ADB76146	ADB76146 Novel hum
45	267	17.4	46	4	AAE13155	AAE13155 Human SCR

ALIGNMENTS

RESULT 1
AAE13167
ID AAE13167 standard; protein; 279 AA.
XX AAE13167;
XX
DT 28-JAN-2002 (first entry)
XX
DE Mouse stem cell growth factor-like protein.
XX

Mouse; stem cell growth factor-like protein; antiinflammatory; nootropic; neuroprotective; vulnary; cytostatic; anticonvulsant; immunostimulant; vasotropic; virucide; dermatological; tranquilliser; cerebroprotective; osteopathic; immunodeficiency syndrome; chronic granulomatous disease; duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS; acquired immune deficiency syndrome; agammaglobulinaemia; thalassemia; Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis; adrenal white matter degeneration; anaemia; neurodegenerative disease; Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID; severe combined immunodeficiency; immune disorder; autoimmune disorder; multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis; autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1; supporting factor for the proliferation of stem cell.

Key	Location/Qualifiers
Peptide	1..21
Protein	22..279
	/label= Signal_peptide
	/note= "Mouse mature stem cell growth factor-like protein"

WO200177169-A2.

18-OCT-2001.

05-APR-2001; 2001WO-US011208.

05-APR-2000; 2000US-00543774.

28-JUN-2000; 2000US-0215733P.

09-JAN-2001; 2001US-00757562.

05-FEB-2001; 2001US-0266614P.

(HYSE-) HYSEQ INC.

(KIRI) KIRIN BEER KK.

Tang TY, Labat I, Tillingbaast JS, Sinku A, Liu C, Drmanac RT;

PI	Stache-Crain B, Dickson M, Mize NK, Nishikawa M;	XX	Stem cell growth factor-like protein; antianemic; anti-HIV; SCR-1;
DR	WPI; 2001-657166/75.	XX	immunostimulant; vulnery; haematopoietic stem cell; gene therapy;
DR	N-PSDB; AAD21727.	XX	supporting factor for proliferation of stem cells; wound healing;
XX		XX	haematopoietic progenitor cell; stromal cell; AIDS; thalassemia;
PT	Novel stem cell growth factor like polypeptides and polynucleotides for	XX	bone marrow transplantation; cord blood transplantation;
PT	identifying modulators useful for treating diseases such as Alzheimer's	XX	chronic granulomatous disease; duplicated immunodeficiency syndrome;
PT	disease, cancer, rheumatoid arthritis, osteoporosis.	XX	agammaglobulinaemia; Wiskott-Aldrich syndrome; haemolytic anaemia;
XX		XX	congenital anaemia; sicklaemia; Gaucher's disease; morphogenesis;
PS	Claim 28; Page 223-224; 232pp; English.	XX	epithelial cell growth; ovarian follicle development; nerve cell growth;
XX		XX	cartilage remodeling; bone growth; immunosuppression; mouse.
XX		XX	Mus musculus.
CC	The patent discloses novel stem cell growth factor-like proteins and	XX	Location/Qualifiers
CC	polynucleotides encoding them. Proteins of the invention are also known	XX	1. .21
CC	as supporting factor for the proliferation of stem cells (SCR-1). Stem	XX	/label= Signal_peptide
CC	cell growth factor-like proteins are useful for supporting proliferation	XX	22. .279
CC	or survival of a stem cell or germ cell which is preferably primordial	XX	/note= "Mature stem cell growth factor-like protein. This
CC	germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem	XX	protein is specifically claimed in claim 9"
CC	cell, The haematopoietic progenitor cell, pluripotent cell or totipotent	XX	
CC	cell. The haematopoietic progenitor cell cultured using stem cell growth	XX	
CC	factor-like proteins can replace as a graft for the bone marrow	XX	
CC	transplantation or cord blood transplantation for treating a variety of	XX	
CC	diseases such as immunodeficiency syndrome, chronic granulomatous	XX	
CC	disease, duplicated immunodeficiency syndrome, agammaglobulinaemia,	XX	
CC	Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),	XX	
CC	thalassaemia, haemolytic anaemia due to enzyme defect, congenital anaemia	XX	
CC	such as sickle cell anaemia, Gaucher's disease, lysosomal storage	XX	
CC	diseases such as mucopolysaccharidosis, adrenal white matter	XX	
CC	degeneration, a variety of cancer and tumours. Proteins of the invention	XX	
CC	are useful for treating diseases such as Parkinson's disease, Alzheimer's	XX	
CC	disease and other neurodegenerative diseases, thrombocytopaenia, immune	XX	
CC	deficiencies and disorders such as severe combined immunodeficiency	XX	
CC	(SCID) and autoimmune disorders such as multiple sclerosis, systemic	XX	
CC	lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary	XX	
CC	inflammation. Sequences of the invention are also useful in gene therapy.	XX	
CC	The present sequence is stem cell growth factor-like protein from mouse	XX	
XX		XX	
SQ	Sequence 279 AA;	XX	
	Query Match 100.0%; Score 1535; DB 4; Length 279;	XX	
	Best Local Similarity 100.0%; Pred. No. 6.2e-109;	XX	
	Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	XX	
Qy	1 MHURLSCPIIFNFMEYIGSQNASRRRRHPNVISQCGGCATCSYNGCLCKPR 60	XX	
Db	1 MHURLSCPIIFNFMEYIGSQNASRRRRHPNVISQCGGCATCSYNGCLCKPR 60	XX	
Qy	61 LFPVLRIGMKQIGVCLSCSPGYGYTRYDINKTKCKVDCTCFNKNFCTKCKSGFYL 120	XX	
Db	61 LFPVLRIGMKQIGVCLSCSPGYGYTRYDINKTKCKVDCTCFNKNFCTKCKSGFYL 120	XX	
Qy	121 HLGKLDSCPEGLEANNHTMECVSIHVCEASESPWSPCKMKGKTGFKRGTRVRDIL 180	XX	
Db	121 HLGKLDSCPEGLEANNHTMECVSIHVCEASESPWSPCKMKGKTGFKRGTRVRDIL 180	XX	
Qy	181 QHPSAKGNLCPPTETFRFCIVORKKCSGERGKGRERKRLKKEKETSSSSDSK 240	XX	
Db	181 QHPSAKGNLCPPTETFRFCIVORKKCSGERGKGRERKRLKKEKETSSSSDSK 240	XX	
Qy	241 GLESSITETPDQNKERQOQKRRARDKQKSVSVTVH 279	XX	
Db	241 GLESSITETPDQNKERQOQKRRARDKQKSVSVTVH 279	XX	
RESULT 2		XX	
ID	ABO44431	XX	
XX	ABO44431 standard; protein; 279 AA.	XX	
AC	ABO44431;	XX	
DT	30-SEP-2003 (first entry)	XX	
XX	Mouse stem cell growth factor-like protein.	XX	
DE		XX	
XX		XX	

Stem cell growth factor-like protein; antianemic; anti-HIV; SCR-1; immunostimulant; vulnery; haematopoietic stem cell; gene therapy; supporting factor for proliferation of stem cells; wound healing; haematopoietic progenitor cell; stromal cell; AIDS; thalassemia; bone marrow transplantation; cord blood transplantation; chronic granulomatous disease; duplicated immunodeficiency syndrome; agammaglobulinaemia; Wiskott-Aldrich syndrome; haemolytic anaemia; congenital anaemia; sicklaemia; Gaucher's disease; morphogenesis; epithelial cell growth; ovarian follicle development; nerve cell growth; cartilage remodeling; bone growth; immunosuppression; mouse.

Mus musculus.

Location/Qualifiers

1. .21

/label= Signal_peptide

22. .279

/note= "Mature stem cell growth factor-like protein. This protein is specifically claimed in claim 9"

US2003044792-A1.

06-MAR-2003.

28-JUN-2001; 2001US-00894912.

28-JUN-2000; 2000US-0215733P.

05-FEB-2001; 2001US-0266614P.

05-APR-2001; 2001US-0282397P.

(TANG/) TANG Y T.

(LABA/) LABAT I.

(DRMA/) DRMANAC R T.

(MIZE/) MIZE N.

(NISH/) NISHIKAWA M.

(CHAO/) CHAO C.

Tang YT, Labat I, Drmanac RT, Mize N, Nishikawa M, Chao C;

WPI; 2003-625403/59.

N-PSDB; ACH04327.

Novel isolated polypeptide having stem cell growth factor activity, useful for promoting wound healing, and as a medicine to proliferate or support human hematopoietic stem cells or human hematopoietic progenitor cells.

Claim 23; Page 80; 96pp; English.

The invention relates to an isolated stem cell growth factor-like polypeptide (referred as supporting factor for proliferation of stem cells (SCR-1)) from mouse or human, or its mature protein portion, or fragment, analogue, variant or derivative, that retains stem cell growth factor activity. Also included are an isolated polynucleotide encoding SCR-1 (or its mature protein portion, or fragment, analogue, variant or derivative, that retains stem cell growth factor activity, or the complement of the polynucleotide), an (expression) vector comprising the SCR-1 polynucleotide, a host cell genetically engineered to contain the SCR-1 polynucleotide in operative association with a regulatory sequence that controls expression of the polynucleotide in the host cell, preparation of the SCR-1 polypeptide, a polypeptide which is an expression product of the SCR-1 polynucleotide (the polypeptide having an activity to support proliferation or survival of haematopoietic stem cell or haematopoietic progenitor cell, with a proviso that C-terminal aa sequence does not comprise the aa sequence appearing as ABO44433), an isolated SCR-1 polypeptide with stem cell growth factor activity and lacking any 10 consecutive aa from ABO44430, an isolated polypeptide with stem cell growth factor activity having at least an aa sequence appearing as ABO44428 and ABO44429, a culture medium comprising the SCR-1 polypeptide to maintain survival of or promote proliferation of a stem cell or germ cell, an anti-SCR-1 antibody, a nucleic acid array comprising the SCR-1 polynucleotide or a unique segment of the SCR-1 polynucleotide attached to a surface, a stromal cell genetically

CC engineered to express the SCR-1 polypeptide to support proliferation or
 CC survival of a stem cell or germ cell and an implant comprising a cell
 CC genetically engineered to express the SCR-1 polypeptide to support
 CC proliferation or survival of a stem cell or germ cell. The SCR-1
 CC polypeptide is useful for identifying a compound that binds to the SCR-1
 CC polypeptide and for maintaining survival of or promoting proliferation of
 CC a stem cell, a germ cell, a haematopoietic stem cell or a haematopoietic
 CC progenitor cell. The SCR-1 polypeptide is useful for promoting wound
 CC healing. The human haematopoietic stem cell or human haematopoietic
 CC progenitor cell culture using the SCR-1 polypeptide can replace as a
 CC graft for the conventional bone marrow transplantation or cord blood
 CC transplantation. The transplantation of haematopoietic stem cells can be
 CC employed as a therapy for treating diseases such as chronic granulomatous
 CC diseases, duplicated immunodeficiency syndrome, agammaglobulinemia,
 CC Wiskott-Aldrich syndrome, AIDS, etc., thalassemia, haemolytic anaemia,
 CC due to enzyme defect, congenital anaemia such as sickle cell disease, Gaucher's
 CC disease etc. The SCR-1 polypeptide is useful for cell growth and
 CC morphogenesis, including tissue specific stem cell growth, epithelial
 CC cell growth and regulation, ovarian follicle development, promoting nerve
 CC cell growth, sustaining neuronal populations, cartilage remodeling, bone
 CC growth and immunosuppression. The present sequence is a mouse SCR-1
 CC protein

XX SQ Sequence 279 AA;

Query Match 100.0%; Score 1535; DB 6; Length 279;
 Best Local Similarity 100.0%; Pred. No. 6.2e-109; Mismatches 0; Indels 0; Gaps 0;
 Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHLRLISCFPIILNFMEYIGSQNASRRRRRHPNVSQGGCGCATCSDYNGCLSCKPR 60
 DB 1 MHLRLISCFPIILNFMEYIGSQNASRRRRRHPNVSQGGCGCATCSDYNGCLSCKPR 60

QY 61 LFFVLERIGMKQIGVCLSSCPSPGYGTRYPDINKCTCKVDCTCFNKNFCTCKSGFYL 120
 DB 61 LFFVLERIGMKQIGVCLSSCPSPGYGTRYPDINKCTCKVDCTCFNKNFCTCKSGFYL 120

QY 121 HLKGLDSCPEGLEANNHTMECVSIHVCEASESWSPQMKKGTGKRGKTETVRDIL 180
 DB 121 HLKGLDSCPEGLEANNHTMECVSIHVCEASESWSPQMKKGTGKRGKTETVRDIL 180

QY 181 QHPSAKGKGNLCPTSETRTCTIVQKCKSGERKGRERKRLKLNKEERKETSSSSDSK 240
 DB 181 QHPSAKGKGNLCPTSETRTCTIVQKCKSGERKGRERKRLKLNKEERKETSSSSDSK 240

QY 241 GLESSIETPDQENKERRQKRRARDKQKQSVSVTVH 279
 DB 241 GLESSIETPDQENKERRQKRRARDKQKQSVSVTVH 279

RESULT 3

AAB99220
 ID AAB99220 standard; protein; 272 AA.

XX AC AAB99220;

XX DT 07-SEP-2001 (first entry)

XX DE Human thrombospondin-30.

XX KW Human; thrombospondin-30; cytostatic; anti-HIV; antiinflammatory;
 KW malignant tumour; haemopathy; HIV infection; immunological disease;
 KW inflammation disease.

XX OS Homo sapiens.

XX PN WO200140294-A1.

XX PD 07-JUN-2001.

XX PF 27-NOV-2000; 2000WO-CN000494.

XX PR 29-NOV-1999; 99CN-00124148.

XX PA (BIOR-) BIORAD GENE DEV LTD SHANGHAI.

XX PI Mao Y, Xie Y;

XX WI WPI; 2001-397948/42.

XX DR N-PSDB; AAH45131.

XX PT Human thrombospondin-30 and polynucleotide is useful in diagnosis and
 PT treatment of, e.g., malignant tumor, hemopathy, HIV infection,
 PT immunological diseases and various inflammatory diseases.

XX PS Claim 1; Page 27-28; 33pp; Chinese.

XX CC The present sequence is the protein sequence for human thrombospondin-30.
 CC Thrombospondin-30 protein and coding sequence are useful in the diagnosis
 CC and treatment of malignant tumor, haemopathy, HIV infection,
 CC immunological diseases and various inflammation diseases. In addition,
 CC thrombospondin-30 protein may be used for screening mimics, agonists,
 CC antagonists or inhibitors, or for use in peptide fingerprinting.
 CC identification. The thrombospondin-30 coding sequence may be used as
 CC primers for nucleic acid amplification reaction or as probes for
 CC hybridisation reaction, or in producing gene chips or microarrays

XX SQ Sequence 272 AA;

Query Match 85.7%; Score 1315.5; DB 4; Length 272;
 Best Local Similarity 87.1%; Pred. No. 3.1e-92;
 Matches 243; Conservative 11; Mismatches 18; Indels 7; Gaps 4;

QY 1 MHLRLISCFPIILNFMEYIGSQNASRRRRRHPNVSQGGCGCATCSDYNGCLSCKPR 60
 DB 1 MHLRLISCFPIILNFMEYIGSQNASRRRRRHPNVSQGGCGCATCSDYNGCLSCKPR 60

QY 61 LFFVLERIGMKQIGVCLSSCPSPGYGTRYPDINKCTCKVDCTCFNKNFCTCKSGFYL 120
 DB 61 LFFVLERIGMKQIGVCLSSCPSPGYGTRYPDINKCTCKVDCTCFNKNFCTCKSGFYL 120

QY 121 HLKGLDSCPEGLEANNHTMECVSIHVCEASESWSPQMKKGTGKRGKTETVRDIL 180
 DB 121 HLKGLDSCPEGLEANNHTMECVSIHVCEASESWSPQMKKGTGKRGKTETVRDIL 180

QY 181 QHPSAKGKGNLCPTSETRTCTIVQKCKSGERKGRERKRLKLNKEERKETSSSSDSK 240
 DB 181 QHPSA--KGNLCPTTNETCTVQKCKSGERKGRERKRLKLNKEERKETS--AIPDSK 236

QY 241 GLESSIETPDQENKERRQKRRARDKQKQSVSVTVH 279
 DB 237 SLESSKEIPEQENK--QQQKRRVQDK-QKSVSVTVH 272

RESULT 4

AAM78328

ID AAM78328 standard; protein; 272 AA.

XX AC AAM78328;

XX DT 06-NOV-2001 (first entry)

XX DE Human protein SEQ ID NO 990.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200157190-A2.

XX PD 09-AUG-2001.

XX PF 05-FEB-2001; 2001WO-US004098.

XX 03-FEB-2000; 2000US-00496914.
 PR 27-APR-2000; 2000US-00560875.
 PR 19-JUN-2000; 2000US-00598075.
 PR 19-JUL-2000; 2000US-00620325.
 PR 01-SEP-2000; 2000US-00654936.
 PR 15-SEP-2000; 2000US-00663561.
 PR 20-OCT-2000; 2000US-00693325.
 PR 30-NOV-2000; 2000US-00728422.
 XX (HYSE-) HYSEQ INC.
 PA Tang YT, Liu CA, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrnan T, Goodrich R;
 XX
 DR WPI; 2001-476283/51.
 DR N-PSDB; AAK51461.
 XX
 XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.
 XX
 XX Claim 20; Page 3214-3215; 6221pp; English.
 XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAW78323-AAW80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 CC (AAK52582) and 3656 (AAW80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication
 XX
 XX Sequence 272 AA;
 SQ
 Query Match 85.7%; Score 1315.5; DB 4; Length 272;
 Best Local Similarity 87.1%; Pred. NO. 3.1e-92;
 Matches 243; Conservative 11; Mismatches 18; Indels 7; Gaps 4;
 QY 1 MHLRLISCFITILFMFYIGSQNASRGRRORRHHNVSGCGGCATCSDYNGCLSKCRP 60
 DB 1 MHLRLISMLFITILFMFYIGSQNASRGRRORRHHNVSGCGGCATCSDYNGCLSKCRP 60
 QY 61 LPFVLRIQKQIGVCLSSCPGYGYTRYPDINKCTKCKVDCTCFNKNFCTKCKSGPYL 120
 DB 61 LPFVLRIQKQIGVCLSSCPGYGYTRYPDINKCTKCKVDCTCFNKNFCTKCKSGPYL 120
 QY 121 HLGKCLDSCPEGLNHNTHMBCVSIHVCRASNSPMSPCMKGKTCGPKRGTTETRVRI 180
 DB 121 HLGKCLDNCPEGLNHNTHMBCVSIHVCRSENNPMSPTCKGTCGPKRGTTETRVREII 180
 QY 181 QHPKAGKNCLEPPTSETCTIVQKCKSGERGKGRERKKNKEKETSSSSDSK 240
 DB 181 QHPSA--KGLCPPTNETKCTVQRKKCKGGERGKGRERKKNKEKETSSSSDSK 240
 QY 241 GLESIETTPDQENKRCQQQKRRARDKQKSVSVSTVH 279
 DB 237 SLESSKEIPEQRNK--QQQKKEVQDK-QKSVSVSTVH 272
 RESULT 5
 ID AAB13168
 XX AAB13168 standard; protein; 272 AA.
 AC AAE13168;
 XX
 XX 28-JAN-2002 (first entry)
 DT
 XX

DE Human stem cell growth factor-like protein #4.
 XX
 KW Human; stem cell growth factor-like protein; antiinflammatory; neutropic;
 KW neuroprotective; vulnerary; cyostatic; anticonvulsant; immunostimulant;
 KW vasotropic; virucide; dermatological; tranquilliser; cerebroprotective;
 KW osteopathic; immunodeficiency syndrome; chronic granulomatous disease;
 KW duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;
 KW acquired immune deficiency syndrome; agammaglobulinaemia; thalassaemia;
 KW Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;
 KW adrenal white matter degeneration; anaemia; neurodegenerative disease;
 KW Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;
 KW severe combined immunodeficiency; immune disorder; autoimmune disorder;
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
 KW autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;
 XX supporting factor for the proliferation of stem cell.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..21
 FT Protein /label= Signal_peptide
 FT 22..272
 FT /note= "Human mature stem cell growth factor-like
 FT protein"
 XX
 XX WO200177169-A2.
 XX 18-OCT-2001.
 XX 05-APR-2001; 2001WO-US011208.
 XX 05-APR-2000; 2000US-00543774.
 XX 28-JUN-2000; 2000US-0215733P.
 XX 09-JAN-2001; 2001US-00757562.
 XX 05-FEB-2001; 2001US-0266614P.
 XX (HYSE-) HYSEQ INC.
 PA (KIRI) KIRIN BEER KK.
 XX
 PI Tang TY, Labat I, Tillinghast JS, Sinku A, Liu C, Drmanac RT;
 PI Stache-Crain B, Dickson M, Mize NK, Nishikawa M;
 XX WPI; 2001-657166/75.
 DR N-PSDB; AAD21728.
 XX
 XX Novel stem cell growth factor like polypeptides and polynucleotides for
 PT identifying modulators useful for treating diseases such as Alzheimer's
 PT disease, cancer, rheumatoid arthritis, osteoporosis.
 PS
 PS Claim 28; Page 226-227; 232pp; English.
 XX
 CC The patent discloses novel stem cell growth factor-like proteins and
 CC polynucleotides encoding them. Proteins of the invention are also known
 CC as supporting factor for the proliferation of stem cells (SCR-1). Stem
 CC cell growth factor-like proteins are useful for supporting proliferation
 CC or survival of a stem cell or germ cell which is preferably primordial
 CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem
 CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent
 CC cell. The haematopoietic progenitor cell cultured using stem cell growth
 CC factor-like proteins can replace as a graft for the bone marrow
 CC transplantation or cord blood transplantation for treating a variety of
 CC diseases such as immunodeficiency syndrome, chronic granulomatous
 CC disease, duplicated immunodeficiency syndrome, agammaglobulinaemia,
 CC Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),
 CC thalassaemia, haemolytic anaemia due to enzyme defect, congenital anaemia
 CC such as sickle cell anaemia, Gaucher's disease, lysosomal storage
 CC diseases such as mucopolysaccharidosis, adrenal white matter
 CC degeneration, a variety of cancer and tumour. Proteins of the invention
 CC are useful for treating diseases such as Parkinson's disease, Alzheimer's
 CC disease and other neurodegenerative diseases, thrombocytopaenia, immune
 CC deficiencies and disorders such as severe combined immunodeficiency
 CC (SCID) and autoimmune disorders such as multiple sclerosis, systemic
 CC lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary

CC inflammation. Sequences of the invention are also useful in gene therapy.
CC The present sequence is stem cell growth factor-like protein from human
XX
SQ Sequence 272 AA;
Query Match 85.7%; Score 1315.5; DB 4; Length 272;
Best Local Similarity 87.1%; Pred. No. 3.1e-92;
Matches 243; Conservative 11; Mismatches 18; Indels 7; Gaps 4;
QY 1 MHLRLISCFPIILNFMWEYIGSQNASRGRQRHHPNVSQCGGCATCSDYNGCLSKPR 60
DB 1 MHLRLISLWFLIILNFMWEYIGSQNASRGRQRHHPNVSQCGGCATCSDYNGCLSKPR 60
QY 61 LFFVLERIGMKQIGVCLSSCPSPGYGTRYPDINKTKCKVDCTCFNKNFCTCKSGFYL 120
DB 61 LFFALERIGMKQIGVCLSSCPSPGYGTRYPDINKTKCKADCDTCFNKNFCTCKSGFYL 120
QY 121 HLKCLDSCPEGLEANNHTMBCVSIHVCEASESPWSPCKMKGKTCGFRGRTTRVRDIL 180
DB 121 HLKCLDNCPEGLEANNHTMBCVSIHVCEASESPWSPCTKKGTCGFRGRTTRVRRII 180
QY 181 QHPSAKGKGNLCPTTSETRTCTIVQRKCKSGRGRKGRERKRLKLNKEERKETSSSDSK 240
DB 181 QHPSA--KGNLCPTTNETRKTCTVQRKCKSGRGRKGRERKRLKLNKEERKETSSSDSK 240
QY 241 GLESSIETPDQENKERQOQKERRARDKQKQSVSVSTVH 279
DB 237 SLESSKEIPEQRENK--QOQKGRKVQDK-QKQSVSVSTVH 272

RESULT 6
AAE13150
ID AAE13150 standard; protein; 272 AA.

AC AAE13150;

DT 28-JAN-2002 (first entry)

DE Human stem cell growth factor-like protein #2.

KW Human; stem cell growth factor-like protein; antiinflammatory; neurotropic;
KW neuroprotective; vulnerary; cytostatic; anticonvulsant; immunostimulant;
KW vasotropic; virucide; dermatological; tranquilliser; cerebroprotective;
KW osteopathic; immunodeficiency syndrome; chronic granulomatous disease;
KW duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;
KW acquired immune deficiency syndrome; agammaglobulinaemia; thalassaemia;
KW Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;
KW adrenal white matter degeneration; anaemia; neurodegenerative disease;
KW Parkinson's disease; Alzheimer's disease; immune disorder; SCID;
KW severe combined immunodeficiency; immune disorder; autoimmune disorder;
KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
KW autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;
KW supporting factor for the proliferation of stem cell.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..21

FT Protein /label= Signal_peptide

FT /note= "Human mature stem cell growth factor-like

protein"

XX WO200177169-A2.

XX 18-OCT-2001.

XX 05-APR-2001; 2001WO-US011208.

XX 05-APR-2000; 2000US-00543774.

XX 28-JUN-2000; 2000US-0215733P.

XX 09-JAN-2001; 2001US-00757562.

XX 05-FEB-2001; 2001US-0266614P.

XX
PA

XX (HYSB-) HYSBQ INC.

XX (KIRI) KIRIN BEER KK.

XX Tang TY, Labat I, Tillinghaast JS, Sinku A, Liu C, Drmanac RT;

XX Stache-Crain B, Dickson M, Mize NK, Nishikawa M;

XX WPI; 2001-657166/75.

XX N-PSDB; AAD21724.

XX Novel stem cell growth factor like polypeptides and polynucleotides for
PT identifying modulators useful for treating diseases such as Alzheimer's
PT disease, cancer, rheumatoid arthritis, osteoporosis.

XX Claim 28; Page 211-212; 232pp; English.

CC The patent discloses novel stem cell growth factor-like proteins and
CC polynucleotides encoding them. Proteins of the invention are also known
CC as supporting factor for the proliferation of stem cells (SCR-1). Stem
CC cell growth factor-like proteins are useful for supporting proliferation
CC or survival of a stem cell or germ cell which is preferably primordial
CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem
CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent
CC cell. The haematopoietic progenitor cell cultured using stem cell growth
CC factor-like proteins can replace as a graft for the bone marrow
CC transplantation or cord blood transplantation for treating a variety of
CC diseases such as immunodeficiency syndrome, chronic granulomatous
CC disease, duplicated immunodeficiency syndrome, agammaglobulinaemia,
CC Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),
CC thalassaemia, haemolytic anaemia due to enzyme defect, congenital anaemia
CC such as sickle cell anaemia, Gaucher's disease, lysosomal storage
CC diseases such as mucopolysaccharidosis, adrenal white matter
CC degeneration, a variety of cancer and tumours. Proteins of the invention
CC are useful for treating diseases such as Parkinson's disease, Alzheimer's
CC disease and other neurodegenerative diseases, thrombocytopaenia, immune
CC deficiencies and disorders such as severe combined immunodeficiency
CC (SCID) and autoimmune disorders such as multiple sclerosis, systemic
CC lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary
CC inflammation. Sequences of the invention are also useful in gene therapy.
CC The present sequence is stem cell growth factor-like protein from human

XX Sequence 272 AA;

Query Match 85.7%; Score 1315.5; DB 4; Length 272;

Best Local Similarity 87.1%; Pred. No. 3.1e-92;

Matches 243; Conservative 11; Mismatches 18; Indels 7; Gaps 4;

QY 1 MHLRLISCFPIILNFMWEYIGSQNASRGRQRHHPNVSQCGGCATCSDYNGCLSKPR 60

DB 1 MHLRLISLWFLIILNFMWEYIGSQNASRGRQRHHPNVSQCGGCATCSDYNGCLSKPR 60

QY 61 LFFVLERIGMKQIGVCLSSCPSPGYGTRYPDINKTKCKVDCTCFNKNFCTCKSGFYL 120

DB 61 LFFALERIGMKQIGVCLSSCPSPGYGTRYPDINKTKCKADCDTCFNKNFCTCKSGFYL 120

QY 121 HLKCLDSCPEGLEANNHTMBCVSIHVCEASESPWSPCKMKGKTCGFRGRTTRVRDIL 180

DB 121 HLKCLDNCPEGLEANNHTMBCVSIHVCEASESPWSPCTKKGTCGFRGRTTRVRRII 180

QY 181 QHPSAKGKGNLCPTTSETRTCTIVQRKCKSGRGRKGRERKRLKLNKEERKETSSSDSK 240

DB 181 QHPSA--KGNLCPTTNETRKTCTVQRKCKSGRGRKGRERKRLKLNKEERKETSSSDSK 240

QY 241 GLESSIETPDQENKERQOQKERRARDKQKQSVSVSTVH 279

DB 237 SLESSKEIPEQRENK--QOQKGRKVQDK-QKQSVSVSTVH 272

RESULT 7

ABR62112

ID ABR62112 standard; protein; 272 AA.

XX

AC ABR62112;

XX 18-AUG-2003 (first entry)
 XX Human clone 1 thrombospondin protein #23.
 XX Human; secreted; stem cell growth factor; cytostatic; haemostatic; neuroprotective; immunostimulant; leukaemia; haemophilia; cancer; degenerative disease; Alzheimer's disease; food supplement; immunological disorder; thrombospondin.
 OS Homo sapiens.
 XX WO2003029405-A2.
 XX 10-APR-2003.
 XX 30-AUG-2002; 2002WO-US027746.
 XX 30-AUG-2001; 2001US-0316368P.
 PR 10-DEC-2001; 2001US-0339739P.
 PR 19-APR-2002; 2002US-00125852.
 XX (HYSB-) HYSEQ INC.
 PA Tang YT;
 PI WPI; 2003-381616/36.
 XX New stem cell growth factor-like polypeptides and polynucleotides, useful for treating e.g. leukemia, hemophilia and degenerative diseases like Alzheimer's disease, and for inducing immune response.
 XX Disclosure; Fig 1; 15pp; English.
 XX The invention relates to new stem cell growth factor-like polypeptides and polynucleotides. The stem cell growth factor-like polypeptides and polynucleotides are useful for inducing differentiation of embryonic and adult stem cells to give rise to different cell types, for treating e.g. leukaemia, haemophilia and degenerative diseases like Alzheimer's disease. They are also useful for generating new tissues and organs that may aid patients in need of transplanted tissues. The polynucleotides are useful as hybridisation probes, oligomers or primers for PCR, for chromosome and gene mapping, in recombinantly producing protein, in generating antisense DNA or RNA, in diagnostics as expressed sequence tags for identifying expressed genes, and for inducing immune response. The polypeptides are useful for generating antibodies that specifically bind the polypeptide, as molecular weight markers, and as a food supplement (e.g. protein or amino acid supplement, and as a carbon, nitrogen or carbohydrate source). Compositions comprising the polypeptides or polynucleotides are useful for the diagnosis, treatment or prevention of cancers, and other immunological disorders. The current sequence represents a human clone 1 thrombospondin protein

XX SQ Sequence 272 AA;
 Query Match 85.7%; Score 1315.5; DB 6; Length 272;
 Best Local Similarity 87.1%; Pred. No. 3.1e-92;
 Matches 243; Conservative 11; Mismatches 18; Indels 7; Gaps 4;

QY 1 MHLRLISCFPIILNFMFYIGSNASRGRRQRMHPNVSQCGGCATCSYNGCLSKPR 60
 DB 1 MHLRLISWLFILNFMFYIGSNASRGRRQRMHPNVSQCGGCATCSYNGCLSKPR 60
 QY 61 LPFVLRIQKQIGVCLSSCPGSGYGYTRYDINKTKCTVDCTCFNKNFCTKCSGYL 120
 DB 61 LPFALERIGKQIGVCLSSCPGSGYGYTRYDINKTKCTKADCDTCFKNFCTKCSGYL 120
 QY 121 HLGKCLDSCPEGLNHNHTMVCIVHCEASESPWSPCKMKGKTCGFKGTETRVADIL 180
 DB 121 HLGKCLDNCPEGLNHNHTMVCIVHCEVSEWNPSPCTCKGKTCGFKGTETRVREII 180
 QY 181 QHPSAKGKGLCPPTSTRCTIVCKKSGRGGKGRGRKPKKLNKPKRKFSSSDSK 240

DB 181 QHPSA--KGNLCPTNETRKTCTVQRKKCKQGRGKKGRGRKRPKNKGRSKB--AIPDSK 236
 QY 241 GLESSIETPDQCKEKERQCKRARDKQCKSVSVSTVH 279
 DB 237 SLESSKIPEQRNK--QCKKRVQDK-QKSVSVSTVH 272

RESULT 8
 ABO44432
 ID ABO44432 standard; protein; 272 AA.
 XX ABO44432;
 XX 30-SEP-2003 (first entry)
 XX Human stem cell growth factor-like protein, SCR 1 #4.
 XX Stem cell growth factor-like protein; antianemic; anti-HIV; SCR-1; immunostimulant; vulnary; haematopoietic stem cell; gene therapy; supporting factor for proliferation of stem cells; wound healing; haematopoietic progenitor cell; stromal cell; AIDS; thalassaemia; bone marrow transplantation; cord blood transplantation;
 KW chronic granulomatous disease; duplicated immunodeficiency syndrome; agammaglobulinaemia; Wiskott-Aldrich syndrome; haemolytic anaemia; congenital anaemia; sicklaemia; Gaucher's disease; morphogenesis; epithelial cell growth; ovarian follicle development; nerve cell growth; cartilage remodeling; bone growth; immunosuppression; human.
 XX Homo sapiens.
 OS
 XX Key
 FH Peptide 1..21 Location/Qualifiers
 FT /label= signal_peptide
 FT Protein 22..272
 FT /note= "Mature stem cell growth factor-like protein. This protein is specifically claimed in claim 9"

US2003044792-A1.
 06-MAR-2003.
 28-JUN-2001; 2001US-00894912.
 28-JUN-2000; 2000US-0215733P.
 05-FEB-2001; 2001US-0266614P.
 05-APR-2001; 2001US-0282397P.
 XX (TANG/) TANG Y T.
 PA (LABA/) LABAT I.
 PA (DRMA/) DRMANAC R T.
 PA (MIZE/) MIZE N.
 PA (NISH/) NISHIKAWA M.
 PA (CHAO/) CHAO C.
 XX Tang YT, Labat I, Drmanac RT, Mize N, Nishikawa M, Chao C;
 XX WPI; 2003-625403/59.
 XX N-PSDB; ACH04328.
 XX Novel isolated polypeptide having stem cell growth factor activity, useful for promoting wound healing, and as a medicine to proliferate or support human hematopoietic stem cells or human hematopoietic progenitor cells.
 XX Claim 23; Page 82; 96pp; English.
 XX The invention relates to an isolated stem cell growth factor-like polypeptide (referred as supporting factor for proliferation of stem cells (SCR-1)) from mouse or human, or its mature protein portion, or fragment, analogue, variant or derivative, that retains stem cell growth factor activity. Also included are an isolated polynucleotide encoding SCR-1 or its mature protein portion, or fragment, analogue, variant or derivative, that retains stem cell growth factor activity, or the

complement of the polynucleotide), an (expression) vector comprising the SCR-1 polynucleotide, a host cell genetically engineered to contain the SCR-1 polynucleotide in operative association with a regulatory sequence that controls expression of the polynucleotide in the host cell, preparation of the SCR-1 polypeptide, a polypeptide which is an expression product of the SCR-1 polynucleotide (the polypeptide having an activity to support proliferation or survival of haematopoietic stem cell or haematopoietic progenitor cell, with a proviso that C-terminal aa sequence does not comprise the aa sequence appearing as ABO44433), an isolated SCR-1 polypeptide with stem cell growth factor activity and lacking any 10 consecutive aas from ABO44430, an isolated polypeptide with stem cell growth factor activity having at least an aa sequence appearing as ABO44428 and ABO44429, a culture medium comprising the SCR-1 polypeptide to maintain survival of or promote proliferation of a stem cell or germ cell, an anti-SCR-1 antibody, a nucleic acid array comprising the SCR-1 polynucleotide or a unique segment of the SCR-1 polynucleotide attached to a surface, a stromal cell genetically engineered to express the SCR-1 polypeptide to support proliferation or survival of a stem cell or germ cell and an implant comprising a cell genetically engineered to express the SCR-1 polypeptide to support proliferation or survival of a stem cell or germ cell. The SCR-1 polypeptide is useful for identifying a compound that binds to the SCR-1 polypeptide and for maintaining survival of or promoting proliferation of a stem cell, a germ cell, a haematopoietic stem cell or a haematopoietic progenitor cell. The SCR-1 polypeptide is useful for promoting wound healing. The human haematopoietic stem cell or human haematopoietic progenitor cell culture using the SCR-1 polypeptide can replace as a graft for the conventional bone marrow transplantation or cord blood transplantation. The transplantation of haematopoietic stem cells can be employed as a therapy for treating diseases such as chronic granulomatous diseases, duplicated immunodeficiency syndrome, agammaglobulinaemia, Wiskott-Aldrich syndrome, AIDS, etc., thalassaemia, haemolytic anaemia due to enzyme defect, congenital anaemia such as sickle cell disease etc. the SCR-1 polypeptide is useful for cell growth and morphogenesis, including tissue specific stem cell growth, epithelial cell growth and regulation, ovarian follicle development, promoting nerve cell growth, sustaining neuronal populations, cartilage remodeling, bone growth and immunosuppression. The present sequence is a Human SCR-1 protein

XX Sequence 272 AA;

Query Match 85.78; Score 1315.5; DB 6; Length 272;
 Best Local Similarity 87.18; Pred. No. 3.1e-92;
 Matches 243; Conservative 11; Mismatches 18; Indels 7; Gaps 4;

QY 1 MHLRLISCFPIILNFMFYIGSQNASRRRRRHPVYSQGGGCATCSYNGCLCKPR 60
 DB 1 MHLRLISLWFIILNFMFYIGSQNASRRRRRHPVYSQGGGCATCSYNGCLCKPR 60
 QY 61 LPFVLERIGMKQIGVCLSSCPGSGYGYTRYPDINKTKKVDCTCFNNFCTCKSGFYL 120
 DB 61 LPFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKTKKADCTCFNNFCTCKSGFYL 120
 QY 121 HLKCLDSCEGLEANNHTWECVSVHCEASENSPWSKKGKTCGFKGTETVRDIL 180
 DB 121 HLKCLDNCPEGLEANNHTWECVSVHCEVSEWNPSPCTCKGKTCGFKGTETVRREII 180
 QY 181 QHPSAKGKGNLCPTSTRTCTIVORKKSGERKGRKRRKRLKNEKERTSSSDSK 240
 DB 181 QHPSA--KGNLCPTNTNCTVORKKQGERKGRKRRKRPKNKESKE--AIPDSK 236
 QY 241 GLSSITPTDQENKRRQOQKRRARDKQKSVSVTVH 279
 DB 237 SLESSKEIPQRENK--QQQKRVQDK-QKSVSVTVH 272

RESULT 9

ABO44413

ID ABO44413 standard; protein; 272 AA.

XX ABO44413;

XX

30-SEP-2003 (first entry)

Human stem cell growth factor-like protein, SCR 1 #1.

Stem cell growth factor-like protein; antianemic; anti-HIV; SCR-1; immunostimulant; vulnary; haematopoietic stem cell; gene therapy; supporting factor for proliferation of stem cells; wound healing; haematopoietic progenitor cell; stromal cell; AIDS; thalassaemia; bone marrow transplantation; cord blood transplantation; chronic granulomatous disease; duplicated immunodeficiency syndrome; agammaglobulinaemia; Wiskott-Aldrich syndrome; haemolytic anaemia; congenital anaemia; sickle cell disease; Gaucher's disease; morphogenesis; epithelial cell growth; ovarian follicle development; nerve cell growth; cartilage remodeling; bone growth; immunosuppression; human.

Homo sapiens.

US2003044792-A1.

06-MAR-2003.

28-JUN-2001; 2001US-00894912.

28-JUN-2000; 2000US-0215733P.

05-FEB-2001; 2001US-0266614P.

05-APR-2001; 2001US-0282337P.

(TANG// TANG Y T.

(LABA// LABAT I.

(DRMA// DRMANAC R T.

(MIZE// MIZE N.

(NISH// NISHIKAWA M.

(CHAO// CHAO C.

Tang YT, Labat I, Drmanac RT, Mize N, Nishikawa M, Chao C;

WPI; 2003-625403/59.

N-PSDB; ACH04323, ACH04324.

Novel isolated polypeptide having stem cell growth factor activity, useful for promoting wound healing, and as a medicine to proliferate or support human hematopoietic stem cells or human hematopoietic progenitor cells.

Claim 23; Fig 3; 96pp; English.

The invention relates to an isolated stem cell growth factor-like polypeptide (referred as supporting factor for proliferation of stem cells (SCR-1)) from mouse or human, or its mature protein portion, or fragment, analogue, variant or derivative, that retains stem cell growth factor activity. Also included are an isolated polynucleotide encoding SCR-1 (or its mature protein portion, or fragment, analogue, variant or derivative, that retains stem cell growth factor activity, or the complement of the polynucleotide), an (expression) vector comprising the SCR-1 polynucleotide, a host cell genetically engineered to contain the SCR-1 polynucleotide in operative association with a regulatory sequence that controls expression of the polynucleotide in the host cell, preparation product of the SCR-1 polypeptide, a polypeptide which is an expression product of the SCR-1 polynucleotide (the polypeptide having an activity to support proliferation or survival of haematopoietic stem cell or haematopoietic progenitor cell, with a proviso that C-terminal aa sequence does not comprise the aa sequence appearing as ABO44433), an isolated SCR-1 polypeptide with stem cell growth factor activity and lacking any 10 consecutive aas from ABO44430, an isolated polypeptide with stem cell growth factor activity having at least an aa sequence appearing as ABO44428 and ABO44429, a culture medium comprising the SCR-1 polypeptide to maintain survival of or promote proliferation of a stem cell or germ cell, an anti-SCR-1 antibody, a nucleic acid array comprising the SCR-1 polynucleotide or a unique segment of the SCR-1 polynucleotide attached to a surface, a stromal cell genetically engineered to express the SCR-1 polypeptide to support proliferation or survival of a stem cell or germ cell and an implant comprising a cell genetically engineered to express the SCR-1 polypeptide to support

CC proliferation or survival of a stem cell or germ cell. The SCR-1
 CC polypeptide is useful for identifying a compound that binds to the SCR-1
 CC polypeptide and for maintaining survival of or promoting proliferation of
 CC a stem cell, a germ cell, a haematopoietic stem cell or a haematopoietic
 CC progenitor cell. The SCR-1 polypeptide is useful for promoting wound
 CC healing. The human haematopoietic stem cell or human haematopoietic
 CC progenitor cell culture using the SCR-1 polypeptide can replace as a
 CC graft for the conventional bone marrow transplantation or cord blood
 CC transplantation. The transplantation of haematopoietic stem cells can be
 CC employed as a therapy for treating diseases such as chronic granulomatous
 CC diseases, duplicated immunodeficiency syndrome, agammaglobulinemia,
 CC Wiskott-Aldrich syndrome, AIDS, etc., thalassemia, haemolytic anaemia,
 CC due to enzyme defect, congenital anaemia such as sicklaemia, Gaucher's
 CC disease etc. The SCR-1 polypeptide is useful for cell growth and
 CC morphogenesis, including tissue specific stem cell growth, epithelial
 CC cell growth and regulation, ovarian follicle development, promoting nerve
 CC cell growth, sustaining neuronal populations, cartilage remodeling, bone
 CC growth and immunosuppression. The present sequence is a Human SCR-1
 CC protein

XX Sequence 272 AA;

Query Match 85.7%; Score 1315.5; DB 6; Length 272;
 Best Local Similarity 87.1%; Pred. No. 3.1e-92;
 Matches 243; Conservative 11; Mismatches 18; Indels 7; Gaps 4;

QY 1 MHRLISCFPIILNFMFYIGSQNASRRGRQRMRHPNVNSQCGGCATCSDYNGCLSCPKR 60
 DB 1 MHRLISWLFILNFMFYIGSQNASRRGRQRMRHPNVNSQCGGCATCSDYNGCLSCPKR 60
 QY 61 LFFVLERIGMKQIGVCLSSPCSGYGYTRYPDINKCTKCVDCDTCFNKNFCTKCKSGFYL 120
 DB 61 LFFALERIGMKQIGVCLSSPCSGYGYTRYPDINKCTKCVDCDTCFNKNFCTKCKSGFYL 120
 QY 121 HLKCLDSCPEGLNANNHMECVSIHVCEASEWSPSPCMKKGKTCGFKGTETRVRDIL 180
 DB 121 HLKCLDNCPEGLNANNHMECVSIHVCEASEWSPSPCTKKGKTCGFKGTETRVREII 180
 QY 181 QHPSAKGKGNLCPTSETTRTCIVORKKCKSGERKGRERKRLKLNKBERKPTSSSDSK 240
 DB 181 QHPSA--KGNLCPTTETRTCTVQRKCKQGERGKGRERKRLKPNKGSKE--AIPDSK 236
 QY 241 GLESSITPDQENKGRQOKERRARDKQKSVSVTVH 279
 DB 237 SLESSKEIPEQRENK--QOKKRVQDK-QKSVSVTVH 272

RESULT 10

AAE13151

ID AAE13151 standard; protein; 273 AA.

XX AAE13151;

AC AAE13151;

XX 28-JAN-2002 (first entry)

XX Human stem cell growth factor-like protein #3.

DE Human; stem cell growth factor-like protein; antiinflammatory; neurotropic;

KW neuroprotective; vulnary; cytostatic; anticonvulsant; immunostimulant;
 KW vasotropic; virucide; dermatological; tranquilliser; cerebroprotective;
 KW osteopathic; immunodeficiency syndrome; chronic granulomatous disease;
 KW duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;
 KW acquired immune deficiency syndrome; agammaglobulinemia; thalassemia;
 KW Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;
 KW adrenal white matter degeneration; anaemia; neurodegenerative disease;
 KW Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;
 KW severe combined immunodeficiency; immune disorder; autoimmune arthritis;
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
 KW autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;
 KW supporting factor for the proliferation of stem cell.

OS Homo sapiens.

XX

PN WO200177169-A2.

XX 18-OCT-2001.

XX 05-APR-2001; 2001WO-US011208.

XX 05-APR-2000; 2000US-00543774.

XX 28-JUN-2000; 2000US-0215733P.

XX 09-JAN-2001; 2001US-00757562.

XX 05-FEB-2001; 2001US-0266614P.

XX (HYSB-) HYSEQ INC.

XX (KIRI) KIRIN BEER KK.

XX Tang TY, Labat I, Tillinghast JS, Sinku A, Liu C, Drmanac RT;

XX Stache-Crain B, Dickson M, Mize NK, Nishikawa M;

XX WPI; 2001-6571166/75.

XX N-PSDB; AAD21725.

XX Novel stem cell growth factor like polypeptides and polynucleotides for

XX identifying modulators useful for treating diseases such as Alzheimer's

XX disease, cancer, rheumatoid arthritis, osteoporosis.

XX Claim 28; Page 214-215; 232pp; English.

XX The patent discloses novel stem cell growth factor-like proteins and

XX polynucleotides encoding them. Proteins of the invention are also known

XX as supporting factor for the proliferation of stem cells (SCR-1). Stem

XX cell growth factor-like proteins are useful for supporting proliferation

XX or survival of a stem cell or germ cell which is preferably primordial

XX germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem

XX cell, haematopoietic progenitor cell, pluripotent cell or totipotent

XX cell. The haematopoietic progenitor cell cultured using stem cell growth

XX factor-like proteins can replace as a graft for the bone marrow

XX transplantation or cord blood transplantation for treating a variety of

XX diseases such as immunodeficiency syndrome, chronic granulomatous

XX disease, duplicated immunodeficiency syndrome, agammaglobulinemia,

XX Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),

XX thalassemia, haemolytic anaemia due to enzyme defect, congenital anaemia

XX such as sickle cell anaemia, Gaucher's disease, lysosomal storage

XX diseases such as mucopolysaccharidosis, adrenal white matter

XX degeneration, a variety of cancer and tumours. Proteins of the invention

XX are useful for treating diseases such as Parkinson's disease, Alzheimer's

XX disease and other neurodegenerative diseases, thrombocytopaenia, immune

XX deficiencies and disorders such as severe combined immunodeficiency

XX (SCID) and autoimmune disorders such as multiple sclerosis, systemic

XX lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary

XX inflammation. Sequences of the invention are also useful in gene therapy.

XX The present sequence is stem cell growth factor-like protein from human

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Query Match 85.4%; Score 1310.5; DB 4; Length 273;

Best Local Similarity 87.1%; Pred. No. 7.5e-92;

Matches 242; Conservative 11; Mismatches 18; Indels 7; Gaps 4;

Sequence 273 AA;

QY 2 HRLISCFPIILNFMFYIGSQNASRRGRQRMRHPNVNSQCGGCATCSDYNGCLSCPKRL 61

DB 3 HRLISWLFILNFMFYIGSQNASRRGRQRMRHPNVNSQCGGCATCSDYNGCLSCPKRL 62

QY 62 FFFVLERIGMKQIGVCLSSPCSGYGYTRYPDINKCTKCVDCDTCFNKNFCTKCKSGFYLH 121

DB 63 FFFALERIGMKQIGVCLSSPCSGYGYTRYPDINKCTKCVDCDTCFNKNFCTKCKSGFYLH 122

QY 122 LGKCLDSCPEGLNANNHMECVSIHVCEASEWSPSPCMKKGKTCGFKGTETRVRDILQ 181

DB 123 LGKCLDNCPEGLNANNHMECVSIHVCEASEWSPSPCTKKGKTCGFKGTETRVREIIQ 182

QY 182 HPSAKGKGNLCPTSETTRTCIVORKKCKSGERKGRERKRLKLNKBERKPTSSSDSKG 241

DB 183 HPSA--KGNLCPTTETRTCTVQRKCKQGERGKGRERKRLKPNKGSKE--AIPDSKS 238

QY 242 LESSIETPDQENKEROQKRRARDKQKSVSVSTVH 279
 DB 239 LESSKEIPEQRENK--QOQKGRVQDK-QKSVSVSTVH 273

RESULT 11
 ABO44414
 ID ABO44414 standard; protein; 273 AA.
 AC ABO44414;
 XX 30-SEP-2003 (first entry)
 DT Human stem cell growth factor-like protein, SCR 1 #2.
 DE
 XX Stem cell growth factor-like protein; antianemic; anti-HIV; SCR-1;
 KW immunostimulant; vulnery; haematopoietic stem cell; gene therapy;
 KW supporting factor for proliferation of stem cells; wound healing;
 KW haematopoietic progenitor cell; stromal cell; AIDS; thalassaemia;
 KW bone marrow transplantation; cord blood transplantation;
 KW chronic granulomatous disease; duplicated immunodeficiency syndrome;
 KW agammaglobulinaemia; Wiskott-Aldrich syndrome; haemolytic anaemia;
 KW congenital anaemia; sicklaemia; Gaucher's disease; morphogenesis;
 KW epithelial cell growth; ovarian follicle development; nerve cell growth;
 KW cartilage remodeling; bone growth; immunosuppression; human.
 XX Homo sapiens.
 OS
 XX US2003044792-A1.
 XX 06-MAR-2003.
 XX 28-JUN-2001; 2001US-00894912.
 XX 28-JUN-2000; 2000US-0215733P.
 PR 05-FEB-2001; 2001US-0266614P.
 PR 05-APR-2001; 2001US-0282397P.
 XX (TANG/) TANG Y T.
 PA (LABA/) LABAT I.
 PA (DRMA/) DRMANAC R T.
 PA (MIZE/) MIZE N.
 PA (NISH/) NISHIKAWA M.
 PA (CHAO/) CHAO C.
 XX Tang YT, Labat I, Drmanac RT, Mize N, Nishikawa M, Chao C;
 PI WPI: 2003-625403/59.
 XX N-PSDB; ACH04325.
 XX Novel isolated polypeptide having stem cell growth factor activity,
 PT useful for promoting wound healing, and as a medicine to proliferate or
 PT support human hematopoietic stem cells or human hematopoietic progenitor
 PT cells.
 XX Claim 23; Page 72-73; 96pp; English.

XX The invention relates to an isolated stem cell growth factor-like
 CC polypeptide (referred as supporting factor for proliferation of stem
 CC cells (SCR-1)) from mouse or human, or its mature protein portion, or
 CC fragment, analogue, variant or derivative, that retains stem cell growth
 CC factor activity. Also included are an isolated polynucleotide encoding
 CC SCR-1 (or its mature protein portion, or fragment, analogue, variant or
 CC derivative, that retains stem cell growth factor activity, or the
 CC complement of the polynucleotide), an (expression) vector comprising the
 CC SCR-1 polynucleotide, a host cell genetically engineered to contain the
 CC SCR-1 polynucleotide in operative association with a regulatory sequence
 CC that controls expression of the polynucleotide in the host cell,
 CC preparation of the SCR-1 polypeptide, a polypeptide which is an
 CC expression product of the SCR-1 polynucleotide (the polypeptide having an
 CC activity to support proliferation or survival of haematopoietic stem cell
 CC or haematopoietic progenitor cell, with a proviso that C-terminal aa
 CC sequence does not comprise the aa sequence appearing as ABO44413), an

CC isolated SCR-1 polypeptide with stem cell growth factor activity and
 CC lacking any 10 consecutive aas from ABO44430, an isolated polypeptide
 CC with stem cell growth factor activity having at least an aa sequence
 CC appearing as ABO44428 and ABO44429, a culture medium comprising of a stem
 CC polypeptide to maintain survival of or promote proliferation of a stem
 CC cell or germ cell, an anti-SCR 1 antibody, a nucleic acid array
 CC comprising the SCR-1 polynucleotide or a unique segment of the SCR-1
 CC polynucleotide attached to a surface, a stromal cell genetically
 CC engineered to express the SCR-1 polypeptide to support proliferation or
 CC survival of a stem cell or germ cell and an implant comprising a cell
 CC genetically engineered to express the SCR-1 polypeptide to support
 CC proliferation or survival of a stem cell or germ cell. The SCR-1
 CC polypeptide is useful for identifying a compound that binds to the SCR-1
 CC polypeptide and for maintaining survival of or promoting proliferation of
 CC a stem cell, a germ cell, a haematopoietic stem cell or a haematopoietic
 CC progenitor cell. The SCR-1 polypeptide is useful for promoting wound
 CC healing. The human haematopoietic stem cell or human haematopoietic
 CC progenitor cell culture using the SCR-1 polypeptide can replace as a
 CC graft for the conventional bone marrow transplantation or cord blood
 CC transplantation. The transplantation of haematopoietic stem cells can be
 CC employed as a therapy for treating diseases such as chronic granulomatous
 CC diseases, duplicated immunodeficiency syndrome, agammaglobulinaemia,
 CC Wiskott-Aldrich syndrome, AIDS, etc., thalassaemia, haemolytic anaemia
 CC due to enzyme defect, congenital anaemia such as sicklaemia, Gaucher's
 CC disease etc. the SCR-1 polypeptide is useful for cell growth and
 CC morphogenesis, including tissue specific stem cell growth, epithelial
 CC cell growth and regulation, ovarian follicle development, promoting nerve
 CC cell growth, sustaining neuronal populations, cartilage remodeling, bone
 CC growth and immunosuppression. The present sequence is a human SCR-1
 CC protein
 XX
 SQ Sequence 273 AA;

Query Match 85.4%; Score 1310.5; DB 6; Length 273;
 Best Local Similarity 87.1%; Pred. No. 7.5e-92;
 Matches 242; Conservative 11; Mismatches 18; Indels 7; Gaps 4;

QY 2 HRLISCFPIILNFMETIGSONASRGRRORRHPNVNVSQCGGCATCSQDYNCLSKCPRL 61
 DB 3 HRLISWLFILNFMETIGSONASRGRRORRHPNVNVSQCGGCATCSQDYNCLSKCPRL 62
 QY 62 PFVLERIGMKQIGVCLSCSPSGYGYTRYPDINKCTCKVDCDTCFNKPFCTKCKSGFYLN 121
 DB 63 PFALERIGMKQIGVCLSCSPSGYGYTRYPDINKCTCKADCDTCFNKPFCTKCKSGFYLN 122
 QY 122 LGKCLDSCPEGLEANNHTMECVSIHVCEASBWSWSPSCMKKGTCTGPKGTTRVADLIQ 181
 DB 123 LGKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPWSPCTKKGKTCGPKGTTRVREIIQ 182
 QY 182 HPSAKGKGLCPPTSETTCTVORCKSKGRRKRRKRLKKEKETSSESSDSKG 241
 DB 183 HPSA--KGNLCPTTNETKCTVQRKKCGKRGKRRKRRKPKNGSKS--AIPDSKS 238
 QY 242 LESSIETPDQENKEROQKRRARDKQKSVSVSTVH 279
 DB 239 LESSKEIPEQRENK--QOQKGRVQDK-QKSVSVSTVH 273

RESULT 12
 AAW85607
 ID AAW85607 standard; protein; 292 AA.
 XX AAW85607;
 AC AAW85607;
 XX 02-MAR-1999 (first entry)
 DT Secreted protein clone da238_6.
 DE
 XX Clone; secreted protein; protein factor; cytokine; lymphokine;
 KW interferon; colony stimulating factor; CSF; interleukin; cloning;
 KW tumour invasion; tumour suppression; immune boosting.
 XX Homo sapiens.
 OS


```

1  MHLRLISWLFILIPMEYIGSQAISRRQRMRHNPVSGCGGCATCSDYNGCLSCRPR 60
61  LPFVLERIGMKQIQIGVCLSSCPSPGYGTYRDPINKCTCKRVDCTCFNKNFCTCKSGFYLL 120
    |||||
61  LPFALERIGMKQIQIGVCLSSCPSPGYGTYRDPINKCTCKKADCTCFNKNFCTCKSGFYLL 120
    |||||
121  HLGHCLDSCPGLRNNHTMBCVSIHVHCEASFWSPSCMKKGKTKCGPKRGTTTAVRDTLL 180
    |||||
121  HLGHCLDNCPEGLRNNHTMBCVSIHVHCEVSEWNPWSCPTKKGTKCGPKRGTTTAVRELI 180
    |||||
181  QHPSAKGKGNLCPTPTSTRCTIVORKKCSKGERGKKRERKRLKLNKERKTKTSSSSDSK 240
    |||||
181  QHPSA--KGNLCPTPTNTRKCTVQRKKCQGERGKKRERKRLKLNKESKZ--AIPDSK 236
    |||||
241  GLESSITFPDQENKEROOQKRRARDKQKQSVSVS 276
    |||||
237  SLESSKPIPEQRNK--OOQKKRVQDKQSGIEVT 270
    |||||

```

RESULT 14	
ABP61846	
ID	ABP61846 standard; protein; 292 AA.
XX	
AC	ABP61846;
XX	
DT	04-OCT-2002 (first entry)
XX	
DE	Human polypeptide SEQ ID NO 200.
XX	
KW	Human; cytostatic; antirheumatic; antiarthritic; vulnary; analgesic;
KW	antiinflammatory; antibacterial; immunosuppressive; antiparkinsonian;
KW	neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;
KW	antitumor; fungicide; antidiabetic; antiasthmatic; antiallergic;
KW	immunostimulant; antiparasitic; secreted protein; transmembrane protein;
KW	cytokine; cell proliferation; cell differentiation; autoimmune disease;
KW	stem cell; growth factor; nervous system disease; neuropathy;
KW	Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW	osteoporosis; severe combined immunodeficiency; SCID; infection;
KW	multiple sclerosis; rheumatoid arthritis; gene therapy.

OS	Homo sapiens.	
XX	US2002065394-A1.	
XX		
PN		
XX	30-MAY-2002.	
PD		
XX		
XX	22-DEC-2000; 2000US-00745763.	
PP		
XX		
XX	18-MAR-1998; 98US-00040963.	
PR		
XX		
XX	(JACO//) JACOBS K.	
PA	(MCCO//) MCCOY J M.	
PA	(LAVA//) LAVALLIE E R.	
PA	(COLL//) COLLINS-RACIE L A.	
PA	(EVAN//) EVANS C.	
PA	(MERB//) MERBERG D.	
PA	(TREA//) TREACY M.	
PA	(SPAU//) SPAULDING V.	
XX		
XX	Jacobs K, Mccoy JM, Lavallie ER, Collins-Racie LA, Evans C;	
PI	Merberg D, Treacy M, Spaulding V;	
PI		
XX		
XX	WPI; 2002-582343/62.	
DR		
DR	N-PSDB; ABQ92060.	
XX		
PT	Novel secreted or transmembrane protein and polynucleotide encoding the	
PT	protein, useful for diagnosis and treatment of neurological disorders,	
PT	cancer, autoimmune diseases, bone disorders and lung or liver fibrosis.	
XX		
XX	Claim 207; Page 203-204; 284pp; English.	
PS		
XX		
XX	The invention relates to human secreted or transmembrane protein (I),	
CC	their fragments and is encoded by specific complementary deoxyribonucleic	
CC	acid sequences.	

acid (cDNA) inserts (II), where the protein is substantially free from other mammalian proteins. (I) are useful for preventing, treating or ameliorating a medical condition, especially immunological treatment or prevention of tumours. (I) exhibits activity relating to angiogenesis, cytokine, cell proliferation, cell differentiation, anti-inflammatory, stem cell growth factor activity and activin or inhibin-related activities. (I) can be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease, accidental damage or genetic disorders. (I) induces the proliferation of neural cells and regeneration of nerve and brain tissue and is useful for the treatment of central and peripheral nervous system diseases and neuropathies, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis. (I) is involved in chemotactic or chemokinetic activity, regulation of haematopoiesis and is useful for treating myeloid or lymphoid cell disorders, platelet disorders such as thrombocytopenia and for regeneration of bone, cartilage, tendon, ligament and/or nerve tissue growth and in tissue repair, healing of burns, incisions, ulcers, for treating osteoporosis, osteoarthritis, bone degenerative disorders or periodontal disease. (I) is also useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, various immune deficiencies and disorders including severe combined immunodeficiency (SCID), bacterial or fungal infections, autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic reactions and conditions, such as asthma or other respiratory problems. (II) is useful to express recombinant protein, as markers for tissues in which the corresponding protein is preferentially expressed and in gene therapy. The present sequence is that of a polypeptide of the invention

AA
SQ
Sequence 292 AA;

Query Match	84.2%	Score 1293	DB 5	Length 292
Best Local Similarity	85.1%	pred. No. 1.7e-90		
Matches 235	Conservative	14	Mismatches 21	Indels 6
				Gaps 3

[illegible]

BR.SIT.T 15

ABR62114
ID ABR62114 standard: protein: 292 AA.

AC ABR62114:

18-AUG-2003 (first entry)

Human secreted protein clone da 288 6 #25.

AA Human; secreted; stem cell growth factor; cytostatic; haemostatic;
KW neuroprotective; immunostimulant; leukaemia; haemophilia; cancer;
KW degenerative disease; Alzheimer's disease; food supplement;
KW immunological disorder.

XX OS Homo sapiens.

```
XX WO2003029405-A2.
XX
XX 10-APR-2003.
XX
XX 30-AUG-2002; 2002WO-US027746.
XX
XX 30-AUG-2001; 2001US-0316368P.
XX
XX 10-DEC-2001; 2001US-0339739P.
XX
XX 19-APR-2002; 2002US-00125852.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT;
XX
XX WPI; 2003-381616/36.
XX
XX New stem cell growth factor-like polypeptides and polynucleotides, useful
XX for treating e.g. leukemia, hemophilia and degenerative diseases like
XX Alzheimer's disease, and for inducing immune response.
XX
XX Disclosure; Fig 2; 151pp; English.
XX
XX The invention relates to new stem cell growth factor-like polypeptides
XX and polynucleotides. The stem cell growth factor-like polypeptides and
XX polynucleotides are useful for inducing differentiation of embryonic and
XX adult stem cells to give rise to different cell types, for treating e.g.
XX leukaemia, haemophilia and degenerative diseases like Alzheimer's
XX disease. They are also useful for generating new tissues and organs that
XX may aid patients in need of transplanted tissues. The polynucleotides are
XX useful as hybridisation probes, oligomers or primers for PCR, for
XX chromosome and gene mapping, in recombinantly producing protein, in
XX generating antisense DNA or RNA, in diagnostics as expressed sequence
XX tags for identifying expressed genes, and for inducing immune response.
XX The polypeptides are useful for generating antibodies that specifically
XX bind the polypeptide, as molecular weight markers, and as a food
XX supplement (e.g. protein or amino acid supplement, and as a carbon,
XX nitrogen or carbohydrate source). Compositions comprising the
XX polypeptides or polynucleotides are useful for the diagnosis, treatment
XX or prevention of cancers, and other immunological disorders. The current
XX sequence represents a human secreted protein clone da_288_6
XX
XX Sequence 292 AA;
XX
XX Query Match 84.2%; Score 1293; DB 6; Length 292;
XX Best Local Similarity 85.1%; Pred. NO. 1.7e-90;
XX Matches 235; Conservative 14; Mismatches 21; Indels 6; Gaps 3;
XX
XX QY 1 MHLRLISCFPIILNFMEYIGSQNASRGRQRMRHFNVSQCGGCATCSDYNGCLSCCKPR 60
XX DB 1 MHLRLISWLFPIILNFMEYIGSQNASRGRQRMRHFNVSQCGGCATCSDYNGCLSCCKPR 60
XX
XX QY 61 LFPVLERIGMKQIGVCLSCPSSGYGTRYTPDINKTKCKVDCDTCFNKNFCTKCKSGPYL 120
XX DB 61 LFPALERIGMKQIGVCLSCPSSGYGTRYTPDINKTKCKADCDTCFNKNFCTKCKSGPYL 120
XX
XX QY 121 HLGKCLDSCPEGLHANNHTMECVSIHVCESEWSFWSPCMKKGTCTGPKRGTTETVRDIL 180
XX DB 121 HLGKCLDNCPEGLHANNHTMECVSIHVCESEWNPSPCTKKGTCGPKRGTTETVRREII 180
XX
XX QY 181 QHPKARGKNGKLCPPPTSETRTCTVQRKCKSGRGRKGRKKNLKEKKEKTSSSSDSK 240
XX DB 181 QHPKSA--KGNLCPPTNETRKTCTVQRKCKQGERGKGRKKNLKEKKEKKE--AIPDSK 236
XX
XX QY 241 GLESSITPDQENKERQOQKRRARDKQKSVSVS 276
XX DB 237 SLESSKEIPEQRENK--QQQKKRKVDKQKSGIEVT 270
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 16:57:05 ; Search time 14.5595 Seconds
(without alignments)
989.298 Million cell updates/sec

Title: US-09-894-912A-32
Perfect score: 1535
Sequence: 1 MHLRLSCFFILNFWYIG.....QKRRARDKQKSVSVTVH 279

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pcp.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	179	11.7	969	2	US-08-284-941-2
2	179	11.7	969	2	US-08-447-642-2
3	179	11.7	969	3	US-09-236-503-2
4	179	11.7	969	5	PCT-US93-02147A-2
5	155	10.1	288	1	US-08-368-852-15
6	154	10.0	799	2	US-08-525-940-23
7	154	10.0	799	2	US-08-976-838-23
8	154	10.0	881	2	US-08-525-940-21
9	154	10.0	881	2	US-08-976-838-21
10	154	10.0	915	2	US-08-525-940-18
11	154	10.0	915	2	US-08-976-838-18
12	154	10.0	915	4	US-09-214-555B-2
13	154	10.0	915	4	US-09-214-555B-7
14	152	9.9	288	2	US-08-525-940-15
15	152	9.9	288	2	US-08-976-838-15
16	142.5	9.3	568	1	US-07-862-021B-14
17	142.5	9.3	568	5	PCT-US93-0316A-14
18	142.5	9.3	807	1	US-07-862-021B-10
19	142.5	9.3	807	1	US-08-313-288B-10
20	142.5	9.3	807	4	US-09-132-769-5
21	142.5	9.3	807	5	PCT-US93-0316A-10
22	139.5	9.1	366	3	US-08-857-076-103
23	139.5	9.1	486	3	US-08-746-559A-5
24	139.5	9.1	516	3	US-08-746-559A-4
25	139.5	9.1	807	4	US-09-132-769-1
26	139.5	9.1	807	4	US-09-640-173-186
27	139.5	9.1	807	4	US-09-713-550-186

28 139.5 9.1 1367 2 US-08-249-687C-2 Sequence 2, Appli
29 139.5 9.1 1367 2 US-08-625-819-2 Sequence 2, Appli
30 139.5 9.1 1367 3 US-08-746-559A-2 Sequence 2, Appli
31 139.5 9.1 1367 4 US-08-864-641B-18 Sequence 18, Appli
32 139.5 9.1 1367 4 US-09-343-551-2 Sequence 2, Appli
33 139 9.1 2523 1 US-08-185-432-18 Sequence 18, Appli
34 139 9.1 2523 4 US-08-899-232-3 Sequence 3, Appli
35 138.5 9.0 379 4 US-09-907-794A-4 Sequence 4, Appli
36 138.5 9.0 379 4 US-09-905-125A-4 Sequence 4, Appli
37 138.5 9.0 379 4 US-09-902-775A-4 Sequence 4, Appli
38 137 8.9 1068 1 US-08-537-210A-2 Sequence 2, Appli
39 137 8.9 1068 3 US-09-113-825-2 Sequence 2, Appli
40 137 8.9 2556 1 US-08-185-432-17 Sequence 17, Appli
41 137 8.9 2556 4 US-08-899-232-2 Sequence 2, Appli
42 136.5 8.9 807 4 US-09-132-769-3 Sequence 3, Appli
43 135.5 8.8 1404 4 US-10-164-595-78 Sequence 78, Appli
44 135 8.8 299 3 US-09-188-930-332 Sequence 332, App
45 135 8.8 299 4 US-09-312-283C-192 Sequence 192, App

ALIGNMENTS

RESULT 1
US-08-284-941-2
; Sequence 2, Application US/08284941
; Patent No. 5863756
; GENERAL INFORMATION:
; APPLICANT: BARR, PHILIP J
; APPLICANT: KIEFER, MICHAEL C
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND
; TITLE OF INVENTION: PACE 4.1 GENE AND POLYPEPTIDES IN CELLS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
; STREET: FIVE PALO ALTO SQUARE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/284,941
; FILING DATE: 2 August 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: NEELEY PH.D., RICHARD L.
; REGISTRATION NUMBER: 30092
; REFERENCE/DOCKET NUMBER: CHIR-009/01US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 843-5070
; TELEFAX: (415) 857-0663
; TELEX: 380816 COOLEY PA
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 969 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-284-941-2

Query Match 11.7%; Score 179; DB 2; Length 969;
Best Local Similarity 23.3%; Pred. No. 1.2e-07;
Matches 56; Conservative 23; Mismatches 89; Indels 72; Gaps 12;
QY 18 YIGSQNASRRRRRRHHFVNVSCQCGGATCSD--YNGCLSKRPLFPVLERIGHKQIGV 75
Db 737 YFGDTAARRRR-----CHKGCTCSRAATQCLSCR-RGPF-----HHQKNT 779

```

QY 76 CLSSCPGYYGTRYPDINKCTCKVDCTCFNK-NFCTKCKSGFYHLGKLDSCPEGL 134
Db 780 CVTLCPAGFYADE--SQRNCKLKHPCKCKVDPEKCTVCKGFSLARGSCIPDCBPGTY 837
QY 135 ANNHTEC-----VSIVHCEAS-----EWS-----PWSP 158
Db 838 FDSELRGCGECHTGTGCVGGRBECIHCAKNFHFHDMKVCVPACGEGFYPEEMPGLPHKV 897
QY 159 CMKKGKTC-----CFKRG-----TETVRDILQHPKSAKGKGNLCPTSETRTCTIVORK 206
Db 898 CRCDCENCLSCAGSSRNCCKTGFTQGTGTCITNHTCSNADETFCEMVKSNRLC--ERK 955

```

RESULT 2

```

US-08-447-642-2
; Sequence 2, Application US/08447642
; Patent No. 5989890
; GENERAL INFORMATION:
; APPLICANT: BARR, PHILIP J
; APPLICANT: KIEFER, MICHAEL C
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND
; TITLE OF INVENTION: PACE 4.1 GENE AND POLYPEPTIDES IN CELLS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
; STREET: FIVE PALO ALTO SQUARE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,642
; FILING DATE: 23-MAY-1995
; CLASSIFICATION: 424
; PRIORITY INFORMATION:
; APPLICATION NUMBER: US 08/284,941
; FILING DATE: 2 August 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: NEELEY PH.D., RICHARD L.
; REGISTRATION NUMBER: 30092
; REFERENCE/DOCKET NUMBER: CHIR-009/01US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 843-5070
; TELEFAX: (415) 857-0663
; TELEX: 380816 COOLEY PA
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 969 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-447-642-2

```

```

Query Match 11.7%; Score 179; DB 2; Length 969;
Best Local Similarity 23.3%; Pred. No. 1.2e-07;
Matches 56; Conservative 23; Mismatches 89; Indels 72; Gaps 12;

QY 18 YIGSQNASRGRQRMRHPNVSOQCGGCATCSD--YNGCLSKPRLFPVLERIGMKQIGV 75
Db 737 YFGDTAARRCRR-----CHKGCTCSSRAATQCLSCR-RGFY-----HHQEMNT 779
QY 76 CLSSCPGYYGTRYPDINKCTCKVDCTCFNK-NFCTKCKSGFYHLGKLDSCPEGL 134
Db 780 CVTLCPAGFYADE--SQRNCKLKHPCKCKVDPEKCTVCKGFSLARGSCIPDCBPGTY 837
QY 135 ANNHTEC-----VSIVHCEAS-----EWS-----PWSP 158
Db 838 FDSELRGCGECHTGTGCVGGRBECIHCAKNFHFHDMKVCVPACGEGFYPEEMPGLPHKV 897

```

```

QY 159 CMKKGKTC-----CFKRG-----TETVRDILQHPKSAKGKGNLCPTSETRTCTIVORK 206
Db 898 CRCDCENCLSCAGSSRNCCKTGFTQGTGTCITNHTCSNADETFCEMVKSNRLC--ERK 955

```

RESULT 3

```

US-09-236-503-2
; Sequence 2, Application US/09236503
; Patent No. 6277590
; GENERAL INFORMATION:
; APPLICANT: BARR, PHILIP J
; APPLICANT: KIEFER, MICHAEL C
; TITLE OF INVENTION: Compositions and Methods for PACE 4 and 4.1 Gene and
; TITLE OF INVENTION: Polypeptides in Cells
; FILE REFERENCE: CHIR-009/04US
; CURRENT APPLICATION NUMBER: US/09/236,503
; EARLIER FILING DATE: 1999-01-25
; EARLIER APPLICATION NUMBER: 08/447,642
; EARLIER FILING DATE: 1995-05-23
; EARLIER APPLICATION NUMBER: 08/284,941
; EARLIER FILING DATE: 1994-08-02
; EARLIER APPLICATION NUMBER: 07/848,629
; EARLIER FILING DATE: 1992-03-09
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 969
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-236-503-2

```

```

Query Match 11.7%; Score 179; DB 3; Length 969;
Best Local Similarity 23.3%; Pred. No. 1.2e-07;
Matches 56; Conservative 23; Mismatches 89; Indels 72; Gaps 12;

QY 18 YIGSQNASRGRQRMRHPNVSOQCGGCATCSD--YNGCLSKPRLFPVLERIGMKQIGV 75
Db 737 YFGDTAARRCRR-----CHKGCTCSSRAATQCLSCR-RGFY-----HHQEMNT 779
QY 76 CLSSCPGYYGTRYPDINKCTCKVDCTCFNK-NFCTKCKSGFYHLGKLDSCPEGL 134
Db 780 CVTLCPAGFYADE--SQRNCKLKHPCKCKVDPEKCTVCKGFSLARGSCIPDCBPGTY 837
QY 135 ANNHTEC-----VSIVHCEAS-----EWS-----PWSP 158
Db 838 FDSELRGCGECHTGTGCVGGRBECIHCAKNFHFHDMKVCVPACGEGFYPEEMPGLPHKV 897
QY 159 CMKKGKTC-----CFKRG-----TETVRDILQHPKSAKGKGNLCPTSETRTCTIVORK 206
Db 898 CRCDCENCLSCAGSSRNCCKTGFTQGTGTCITNHTCSNADETFCEMVKSNRLC--ERK 955

```

RESULT 4

```

PCT-US93-02147A-2
; Sequence 2, Application PC/TUS9302147A
; GENERAL INFORMATION:
; APPLICANT: BARR, PHILIP J
; APPLICANT: KIEFER, MICHAEL C
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND
; TITLE OF INVENTION: PACE 4.1 GENE AND POLYPEPTIDES IN CELLS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
; STREET: FIVE PALO ALTO SQUARE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/02147A
; FILING DATE: 19930309
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,629
; FILING DATE: 09-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: NESLEY PH.D., RICHARD L
; REGISTRATION NUMBER: 30092
; REFERENCE/DOCKET NUMBER: CHIR-009/00US
; TELEPHONE: (415) 494-7622
; TELEFAX: (415) 857-0663
; TELEX: 380816 COOLEY PA
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 969 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US93-02147A-2

Query Match 11.7% Score 179; DB 5; Length 969;
Best Local Similarity 23.3%; Pred. No. 1.2e-07;
Matches 56; Conservative 23; Mismatches 89; Indels 72; Gaps 12;

QY 18 YIGSONASRRQRMRHPNVSOQCGCATCSG--YNGCLSCPKPRLFFVLIRIGMKQIGV 75
DB 737 YFGDTAARCR-----CHKGCETSSRAATQCLSCR-RGFY-----HHQEMNT 779
QY 76 CLSCSPSGYGTYPDINKTKKVDCTCPNK-NFTCKKSGFYHLHGKLDSCSPRGLE 134
DB 780 CVTLCPAGPYADE--SQKNCLKCHPSCKKVDPEKCTVCKEGFSLARGSCIPDCERGTY 837
QY 135 ANNHTWEC-----VSIHVCEAS-----EWS-----PWSP 158
DB 838 FDELIRGCEHTTCTGTCVPGREBCHCAKHFHDMKCVACGEGFYPEEMFGLPHKV 897
QY 159 CMKKGKTC-----GFKRG-----TETVRDILQHPSAKGNLCPPTSETRTCIQVRK 206
DB 898 CRRCDENCLSCAGSRRNCRKGTGTQGTGTCITNHTCSNADETFCBMVKSRLC--ERK 955

RESULT 5
US-08-368-852-15
; Sequence 15, Application US/08368852
; Patent No. 5691183
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; APPLICANT: Miranda, Luis R.
; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES
; TITLE OF INVENTION: ENCODING SAID PROTEASES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: 15
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/368,852
; FILING DATE: 05-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Verber, Carol Talkington
```

```
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: 2848-11
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/863-9700
; TELEFAX: 303/863-0223
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 288 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-368-852-15

Query Match 10.1% Score 155; DB 1; Length 288;
Best Local Similarity 28.2%; Pred. No. 4e-06;
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;

QY 35 PNVSQ-GCOG-GCATCSDYNGCLSCPKPRLFFVLIRIGMKQIGVCLSCSPSGFYGTYPDI 92
DB 13 PECSEVGCDDGPDHCND---CL---HYYYKLK---NNTRICVSSCPPGHY---HADK 58
QY 93 NKCTKCKVDCTCPNK--NFTCKKSGFYHL--LKGKLDSCSPBGLRANNHTWECVSIHVC 148
DB 59 KRCRKCAPNCSECFSGSHGDCMCKYGYFLNEETNSCVTHCPDGSYQDTKKNLG----- 112
QY 149 EASWSPWSPCMKKGKTC-GFKRGCTETVRDILQHPSAKGNLCPPTSB 197
DB 113 -----RKCSNKKTCFTBFHKCTCRDGLSLQ-----GSRCSVSCB 147

RESULT 6
US-08-525-940-23
; Sequence 23, Application US/08525940
; Patent No. 5866351
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; APPLICANT: Miranda, Luis R.
; APPLICANT: Wolf, Joseph R.
; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES
; TITLE OF INVENTION: ENCODING SAID PROTEASES
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,940
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/368,852
; FILING DATE: 01-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/088,322
; FILING DATE: 07-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2848-11-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 799 amino acids
```

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-525-940-23

Query Match 10.0%; Score 154; DB 2; Length 799;

Best Local Similarity 28.2%; Pred. No. 1.6e-05;
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;

Qy 35 PNVSQ-GCOG-GCATCSYNGCLSCPKRLPFVLERIGMKQIGVCLSSCPGYYGTRYPDI 92
Db 524 PECSEVCGDGPDPHCND---CL-----HYYYKLR---NNTRICVSSCPPGHY---HADK 569
Qy 93 NKCTKCKVDCTCFNK--NFCTKCKSGFYHLH--LGKCLDSCPEGLANNHTMVCVSIHVC 148
Db 570 KRCRKCAPNCESCFSGSHGDCQMSCKYGYFLNEETNSCVTHCPDGSYQDTKKNLG----- 623
Qy 149 EASEWSPWSPCKMGKTC-GFKRGTTETVRDILQHPKSAKGKGNLCPPTSE 197
Db 624 -----RKCSNCKTCTBFHNTCTCRDGLSLQ-----GSRCSVSCB 658

RESULT 7

US-08-976-838-23
Sequence 23, Application US/08976838
Patent No. 5981259

GENERAL INFORMATION:
APPLICANT: Franzusoff, Alex
TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID
TITLE OF INVENTION: MOLECULES
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.

ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/976,838

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Connell, Gary J.

REGISTRATION NUMBER: 32,020

REFERENCE/DOCKET NUMBER: 2848-11-C2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 863-9700

TELEFAX: (303) 863-0223

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 799 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-976-838-23

Query Match 10.0%; Score 154; DB 2; Length 799;

Best Local Similarity 28.2%; Pred. No. 1.6e-05;
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;

Qy 35 PNVSQ-GCOG-GCATCSYNGCLSCPKRLPFVLERIGMKQIGVCLSSCPGYYGTRYPDI 92
Db 524 PECSEVCGDGPDPHCND---CL-----HYYYKLR---NNTRICVSSCPPGHY---HADK 569
Qy 93 NKCTKCKVDCTCFNK--NFCTKCKSGFYHLH--LGKCLDSCPEGLANNHTMVCVSIHVC 148
Db 570 KRCRKCAPNCESCFSGSHGDCQMSCKYGYFLNEETNSCVTHCPDGSYQDTKKNLG----- 623

Qy 149 EASEWSPWSPCKMGKTC-GFKRGTTETVRDILQHPKSAKGKGNLCPPTSE 197
Db 624 -----RKCSNCKTCTBFHNTCTCRDGLSLQ-----GSRCSVSCB 658

RESULT 8

US-08-525-940-21
Sequence 21, Application US/08525940
Patent No. 5866351

GENERAL INFORMATION:

APPLICANT: Franzusoff, Alex

APPLICANT: Miranda, Luis R.

APPLICANT: Wolf, Joseph R.

TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES

TITLE OF INVENTION: ENCODING SAID PROTEASES

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross & McIntosh

STREET: 1700 Lincoln Street, Suite 3500

CITY: Denver

STATE: Colorado

COUNTRY: U.S.A.

ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/525,940

FILING DATE:

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/368,852

FILING DATE: 01-JAN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/088,322

FILING DATE: 07-JUL-1993

ATTORNEY/AGENT INFORMATION:

NAME: Connell, Gary J.

REGISTRATION NUMBER: 32,020

REFERENCE/DOCKET NUMBER: 2848-11-C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 863-9700

TELEFAX: (303) 863-0223

INFORMATION FOR SEQ ID NO: 21:

SEQUENCE CHARACTERISTICS:

LENGTH: 881 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-525-940-21

Query Match 10.0%; Score 154; DB 2; Length 881;

Best Local Similarity 28.2%; Pred. No. 1.8e-05;
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;

Qy 35 PNVSQ-GCOG-GCATCSYNGCLSCPKRLPFVLERIGMKQIGVCLSSCPGYYGTRYPDI 92
Db 606 PECSEVCGDGPDPHCND---CL-----HYYYKLR---NNTRICVSSCPPGHY---HADK 651
Qy 93 NKCTKCKVDCTCFNK--NFCTKCKSGFYHLH--LGKCLDSCPEGLANNHTMVCVSIHVC 148
Db 652 KRCRKCAPNCESCFSGSHGDCQMSCKYGYFLNEETNSCVTHCPDGSYQDTKKNLG----- 705
Qy 149 EASEWSPWSPCKMGKTC-GFKRGTTETVRDILQHPKSAKGKGNLCPPTSE 197
Db 706 -----RKCSNCKTCTBFHNTCTCRDGLSLQ-----GSRCSVSCB 740

RESULT 9

US-08-976-838-21

; Sequence 21, Application US/08976838
; Patent No. 5981259
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,838
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2848-11-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 881 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-976-838-21

Query Match 10.0%; Score 154; DB 2; Length 881;
Best Local Similarity 28.2%; Pred. No. 1.8e-05;
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;

QY 35 PNVSQ-GCQG-GCATCSYNGCLSCPKLPFLVLERIGMKQIGVCLSSCPGSGYGYTRYDPI 92
Db 606 PECSEVCGDGGPDHND---CL---HYIYKLGK---NNTRICVSSCPPGHY---HADK 651

QY 93 NKCTCKKVDCTCFNK--NFTCKCKSGFYHL--LGKCLDSCPEGLANNHTMBCVSIHVC 148
Db 652 KRCRKCAPNCSCFGSHGDCMCKYGYFLNEETNSCVTHCPDGSYQDTKKNLG----- 705

QY 149 EASEWSPWSPCKKGTCTGFKRGTTETVRDILQHPSAKGNLCPPTSE 197
Db 706 -----RKCSNCKTCTFHNTCECRDGLSLQ-----GSRCSVSCE 740

RESULT 10
US-08-976-838-18
; Sequence 18, Application US/08525940
; Patent No. 5866351
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; APPLICANT: Miranda, Luis R.
; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES
; TITLE OF INVENTION: ENCODING SAID PROTEASES
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,940
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/368,852
; FILING DATE: 01-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/088,322
; FILING DATE: 07-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2848-11-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-525-940-18

Query Match 10.0%; Score 154; DB 2; Length 915;
Best Local Similarity 28.2%; Pred. No. 1.9e-05;
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;

QY 35 PNVSQ-GCQG-GCATCSYNGCLSCPKLPFLVLERIGMKQIGVCLSSCPGSGYGYTRYDPI 92
Db 640 PECSEVCGDGGPDHND---CL---HYIYKLGK---NNTRICVSSCPPGHY---HADK 685

QY 93 NKCTCKKVDCTCFNK--NFTCKCKSGFYHL--LGKCLDSCPEGLANNHTMBCVSIHVC 148
Db 686 KRCRKCAPNCSCFGSHGDCMCKYGYFLNEETNSCVTHCPDGSYQDTKKNLG----- 739

QY 149 EASEWSPWSPCKKGTCTGFKRGTTETVRDILQHPSAKGNLCPPTSE 197
Db 740 -----RKCSNCKTCTFHNTCECRDGLSLQ-----GSRCSVSCE 774

RESULT 11
US-08-976-838-18
; Sequence 18, Application US/08976838
; Patent No. 5981259
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,838
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2848-11-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-976-838-18

Query Match      10.0%; Score 154; DB 2; Length 915;
Best Local Similarity 28.2%; Pred. No. 1.9e-05;
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;

QY 35 PNVSQ-GCOG-GCATSDYNGCLSCPKRPLFFVLERIGMKQIGVCLSSCPGSGYGTYPDI 92
DB 640 PECSEVGCDDPGDPHCND---CL---HYYYKLG---NNTRICVSSCPGHHY---HADK 685

QY 93 NKCTKCKVDCDTCFNK--NFTCKKSGFYH--LGKCLDSCPGLEANNHTECVSIVHC 148
DB 686 KRCKKCAPNCSCFGSHGDCMSCKYGYFLNEETNSCVTHCPDGSYQDTKKNLC----- 739

QY 149 EASEWSPWSPCKMKGKTC-GFRKGTETRVDRDILQHPSSAKGKGNLCPTSE 197
DB 740 -----RKSENCKTCTEFHNCTECRDGLSLQ-----GSRCSVSC 774

RESULT 12
US-09-214-555B-2
; Sequence 2, Application US/09214555B
; Patent No. 6380171
; GENERAL INFORMATION:
; APPLICANT: INSTITUT DE RECHERCHE CLINIQUE DE MONTRAL
; TITLE OF INVENTION: PRO-PROTEIN CONVERTING ENZYME
; FILE REFERENCE: PRO-PROTEIN CONVER ENZ
; CURRENT APPLICATION NUMBER: US/09/214,555B
; CURRENT FILING DATE: 1999-01-04
; PRIOR APPLICATION NUMBER: 60/021,008
; PRIOR FILING DATE: 1996-07-26
; PRIOR APPLICATION NUMBER: 2,203,745
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-555B-2

Query Match      10.0%; Score 154; DB 4; Length 915;
Best Local Similarity 28.2%; Pred. No. 1.9e-05;
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;

QY 35 PNVSQ-GCOG-GCATSDYNGCLSCPKRPLFFVLERIGMKQIGVCLSSCPGSGYGTYPDI 92
DB 640 PECSEVGCDDPGDPHCND---CL---HYYYKLG---NNTRICVSSCPGHHY---HADK 685

QY 93 NKCTKCKVDCDTCFNK--NFTCKKSGFYH--LGKCLDSCPGLEANNHTECVSIVHC 148
DB 686 KRCKKCAPNCSCFGSHGDCMSCKYGYFLNEETNSCVTHCPDGSYQDTKKNLC----- 739

QY 149 EASEWSPWSPCKMKGKTC-GFRKGTETRVDRDILQHPSSAKGKGNLCPTSE 197
DB 740 -----RKSENCKTCTEFHNCTECRDGLSLQ-----GSRCSVSC 774

RESULT 12
US-09-214-555B-2
; Sequence 2, Application US/09214555B
; Patent No. 6380171
; GENERAL INFORMATION:
; APPLICANT: INSTITUT DE RECHERCHE CLINIQUE DE MONTRAL
; TITLE OF INVENTION: PRO-PROTEIN CONVERTING ENZYME
; FILE REFERENCE: PRO-PROTEIN CONVER ENZ
; CURRENT APPLICATION NUMBER: US/09/214,555B
; CURRENT FILING DATE: 1999-01-04
; PRIOR APPLICATION NUMBER: 60/021,008
; PRIOR FILING DATE: 1996-07-26
; PRIOR APPLICATION NUMBER: 2,203,745
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-555B-2

Query Match      10.0%; Score 154; DB 4; Length 915;
Best Local Similarity 28.2%; Pred. No. 1.9e-05;
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;

QY 35 PNVSQ-GCOG-GCATSDYNGCLSCPKRPLFFVLERIGMKQIGVCLSSCPGSGYGTYPDI 92
DB 640 PECSEVGCDDPGDPHCND---CL---HYYYKLG---NNTRICVSSCPGHHY---HADK 685

QY 93 NKCTKCKVDCDTCFNK--NFTCKKSGFYH--LGKCLDSCPGLEANNHTECVSIVHC 148
DB 686 KRCKKCAPNCSCFGSHGDCMSCKYGYFLNEETNSCVTHCPDGSYQDTKKNLC----- 739

QY 149 EASEWSPWSPCKMKGKTC-GFRKGTETRVDRDILQHPSSAKGKGNLCPTSE 197
DB 740 -----RKSENCKTCTEFHNCTECRDGLSLQ-----GSRCSVSC 774

RESULT 13
US-09-214-555B-7
; Sequence 7, Application US/09214555B
```

```
; Patent No. 6380171
; GENERAL INFORMATION:
; APPLICANT: INSTITUT DE RECHERCHE CLINIQUE DE MONTRAL
; TITLE OF INVENTION: PRO-PROTEIN CONVERTING ENZYME
; FILE REFERENCE: PRO-PROTEIN CONVER ENZ
; CURRENT APPLICATION NUMBER: US/09/214,555B
; CURRENT FILING DATE: 1999-01-04
; PRIOR APPLICATION NUMBER: 60/021,008
; PRIOR FILING DATE: 1996-07-26
; PRIOR APPLICATION NUMBER: 2,203,745
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-555B-7

Query Match      10.0%; Score 154; DB 4; Length 915;
Best Local Similarity 28.2%; Pred. No. 1.9e-05;
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;

QY 35 PNVSQ-GCOG-GCATSDYNGCLSCPKRPLFFVLERIGMKQIGVCLSSCPGSGYGTYPDI 92
DB 640 PECSEVGCDDPGDPHCND---CL---HYYYKLG---NNTRICVSSCPGHHY---HADK 685

QY 93 NKCTKCKVDCDTCFNK--NFTCKKSGFYH--LGKCLDSCPGLEANNHTECVSIVHC 148
DB 686 KRCKKCAPNCSCFGSHGDCMSCKYGYFLNEETNSCVTHCPDGSYQDTKKNLC----- 739

QY 149 EASEWSPWSPCKMKGKTC-GFRKGTETRVDRDILQHPSSAKGKGNLCPTSE 197
DB 740 -----RKSENCKTCTEFHNCTECRDGLSLQ-----GSRCSVSC 774

RESULT 14
US-08-525-940-15
; Sequence 15, Application US/08525940
; Patent No. 5866351
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; APPLICANT: Miranda, Luis R.
; APPLICANT: Wolf, Joseph R.
; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES
; TITLE OF INVENTION: ENCODING SAID PROTEASES
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,940
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/368,852
; FILING DATE: 01-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/088,322
; FILING DATE: 07-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2848-11-C1
```


TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 288 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-525-940-15

Query Match 9.9%; Score 152; DB 2; Length 288;
Best Local Similarity 27.7%; Pred. No. 7.4e-06;
Matches 48; Conservative 24; Mismatches 53; Indels 48; Gaps 12;
QY 35 PNVSQ-GCQG-GCATCSDYNGCLSKPRLFFVLERIGMKQIGVCLSSCPGSGYYGTRYPDI 92
DB 13 PECSEVGCDDGPDHCND---CL---HYVYKLGK---NNTRICVSSCPPGHY---HADK 58
QY 93 NKCTKCKVDCDTCFNK--NFCTKCKSGFYLH--LGKCLDSCPEGLEANNHTMECVSI VHC 148
DB 59 KRCRKCAPNCESCFCGSHGDCMSCKYGVFLNBEETNSCVTHCPDGSYQDTKKNLC----- 112
QY 149 EASEWSPWSPCKMKKGTCTGFRGTE-----TVRDILQHPSAKGNLCPPTSE 197
DB 113 -----RKCSNFKTC-----TEFHICTCRDGLSL-----QGSRCVSCE 147

RESULT 15
US-08-976-838-15
Sequence 15, Application US/08976838
Patent No. 5981259
GENERAL INFORMATION:
APPLICANT: Franzusoff, Alex
TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID
TITLE OF INVENTION: MOLECULES
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/976,838
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2848-11-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 288 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-976-838-15

Query Match 9.9%; Score 152; DB 2; Length 288;
Best Local Similarity 27.7%; Pred. No. 7.4e-06;
Matches 48; Conservative 24; Mismatches 53; Indels 48; Gaps 12;
QY 35 PNVSQ-GCQG-GCATCSDYNGCLSKPRLFFVLERIGMKQIGVCLSSCPGSGYYGTRYPDI 92

DB 13 PECSEVGCDDGPDHCND---CL---HYVYKLGK---NNTRICVSSCPPGHY---HADK 58
QY 93 NKCTKCKVDCDTCFNK--NFCTKCKSGFYLH--LGKCLDSCPEGLEANNHTMECVSI VHC 148
DB 59 KRCRKCAPNCESCFCGSHGDCMSCKYGVFLNBEETNSCVTHCPDGSYQDTKKNLC----- 112
QY 149 EASEWSPWSPCKMKKGTCTGFRGTE-----TVRDILQHPSAKGNLCPPTSE 197
DB 113 -----RKCSNFKTC-----TEFHICTCRDGLSL-----QGSRCVSCE 147

Search completed: June 29, 2004, 17:04:01
Job time : 14.5595 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model
Run on: June 29, 2004, 17:02:26 ; Search time 37.5837 Seconds
(without alignments)
2098.641 Million cell updates/sec

Title: US-09-894-912A-32
Perfect score: 1535
Sequence: 1 MHLRLISCFPIILFMFYIG.....QKGRARDKQKSVSVTVH 279

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1166195 seqs, 282705291 residues

Total number of hits satisfying chosen parameters: 1166195

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1535	100.0	279	10	US-09-894-912A-32
2	1315.5	85.7	272	10	US-09-894-912A-10
3	1315.5	85.7	272	10	US-09-894-912A-34
4	1315.5	85.7	272	14	US-10-125-852-23
5	1310.5	85.4	273	10	US-09-894-912A-13
6	1293	84.2	292	9	US-09-745-763-166
7	1293	84.2	292	10	US-09-894-912A-48
8	1293	84.2	292	14	US-10-125-852-25
9	1285	83.7	265	10	US-09-894-912A-26
10	1222.5	79.6	251	10	US-09-894-912A-16
11	1187	77.3	225	14	US-10-185-770-4
12	1135	73.9	239	12	US-10-087-192-1782
13	1100	71.7	239	15	US-10-094-886-172
14	1045	68.1	195	15	US-10-094-886-176
15	1041	67.8	195	15	US-10-094-886-174

16	1018	66.3	180	12	US-10-087-192-1779
17	759	49.4	160	10	US-09-894-912A-14
18	759	49.4	160	12	US-10-276-774-1744
19	648	42.2	263	14	US-10-125-852-18
20	640.5	41.7	243	14	US-10-125-852-21
21	632	41.0	265	14	US-10-125-852-24
22	629	41.0	229	10	US-09-894-912A-25
23	616.5	40.2	236	15	US-10-108-260A-4829
24	571.5	37.2	243	14	US-10-185-770-2
25	569.5	37.1	243	14	US-10-125-852-13
26	569.5	37.1	243	16	US-10-467-042-12
27	565	36.8	250	14	US-10-125-852-3
28	536	34.9	222	14	US-10-125-852-15
29	531.5	34.6	229	14	US-10-125-852-6
30	403.5	26.3	190	12	US-10-087-192-873
31	370.5	24.1	161	12	US-10-087-192-876
32	267	17.4	46	10	US-09-894-912A-18
33	241.5	15.7	131	14	US-10-125-852-9
34	208	13.6	110	14	US-10-125-852-11
35	193	12.6	37	10	US-09-894-912A-20
36	187	12.2	1548	14	US-10-180-903-2
37	181.5	11.8	43	14	US-10-125-852-7
38	179	11.7	969	10	US-09-961-403-6
39	170.5	11.1	942	15	US-10-369-493-5070
40	150	9.8	42	10	US-09-894-912A-22
41	145.5	9.5	827	15	US-10-369-493-5073
42	144.5	9.4	2471	15	US-10-190-115-27
43	144.5	9.4	2471	15	US-10-369-072-27
44	142.5	9.3	807	14	US-10-205-194-133
45	140.5	9.2	660	12	US-09-939-853A-96

ALIGNMENTS

RESULT 1
US-09-894-912A-32
; Sequence 32, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; FACTOR-LIKE POLYPEPTIDES AND POLYNUCLEOTIDES
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894, 912A
; PRIOR FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266, 614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215, 733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757, 562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543, 774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 32
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-894-912A-32

Query Match 100.0%; Score 1535; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 1.2e-114;
Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 1 MHLRLISCFPIILFMFYIGSQNASRGRRQRMHPNYSQGGCATCDYNGCLSCKPR 60
Db 1 MHLRLISCFPIILFMFYIGSQNASRGRRQRMHPNYSQGGCATCDYNGCLSCKPR 60
Qy 61 LFFVLEIRIGHKQIGVCLSCFSPGGYGYTRYPPDINKCKVDCTCFNKNFCTKCKSGPYL 120

Db 61 LPFVLERIGMKQIGVCLSSCPGSGYGYTRYPDINKTKCKVDCDTCFKNKFCCKSGFYL 120
Qy 121 HLKCLDSCPEGLEANNHTECVSIHVCSEASWSPSCMKKGTCTGFKGTETVRDIL 180
Db 121 HLKCLDSCPEGLEANNHTECVSIHVCSEASWSPSCMKKGTCTGFKGTETVRDIL 180
Qy 181 QHPSAKGKGNLCPPTSETRTCTIVQRKCKSGRGGKGRERKRLKLNKBERKETSSSDSK 240
Db 181 QHPSAKGKGNLCPPTSETRTCTIVQRKCKSGRGGKGRERKRLKLNKBERKETSSSDSK 240
Qy 241 GLESSIETPDQENKRRQOQKRRARDKQKSVSVSTVH 279
Db 241 GLESSIETPDQENKRRQOQKRRARDKQKSVSVSTVH 279

RESULT 2

US-09-894-912A-10
; Sequence 10, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; CURRENT FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Homo sapiens

US-09-894-912A-10

Query Match 85.7%; Score 1315.5; DB 10; Length 272;
Best Local Similarity 87.1%; Pred. No. 3.8e-97;
Matches 243; Conservative 11; Mismatches 18; Indels 7; Gaps 4;

Qy 1 MHLRLISCFPIILNFMFYIGSQNASRGRORRHHPNVSQCGGCATCSDYNGCLSKPR 60
Db 1 MHLRLISWLPILNFMFYIGSQNASRGRORRHHPNVSQCGGCATCSDYNGCLSKPR 60
Qy 61 LPFVLERIGMKQIGVCLSSCPGSGYGYTRYPDINKTKCKVDCDTCFKNKFCCKSGFYL 120
Db 61 LPFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKTKCKADCDTCFKNKFCCKSGFYL 120
Qy 121 HLKCLDSCPEGLEANNHTECVSIHVCSEASWSPSCMKKGTCTGFKGTETVRDIL 180
Db 121 HLKCLDNCPEGLEANNHTECVSIHVCSEWNPSPCTKKGKTCGFKGTETVRREII 180
Qy 181 QHPSAKGKGNLCPPTSETRTCTIVQRKCKSGRGGKGRERKRLKLNKBERKETSSSDSK 240
Db 181 QHPSA--KGNLCPPTNETRKTIVQRKCKQKGRGGKGRERKRLKPNKGSKE--AIPDSK 236
Qy 241 GLESSIETPDQENKRRQOQKRRARDKQKSVSVSTVH 279
Db 237 SLESSKEIPEQRENK--QQKKRKVDK-QKSVSVSTVH 272

RESULT 3

US-09-894-912A-34
; Sequence 34, Application US/09894912A

; Publication No. US20030044792A1

; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; CURRENT FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Homo sapiens

US-09-894-912A-34

Query Match 85.7%; Score 1315.5; DB 10; Length 272;
Best Local Similarity 87.1%; Pred. No. 3.8e-97;
Matches 243; Conservative 11; Mismatches 18; Indels 7; Gaps 4;

Qy 1 MHLRLISCFPIILNFMFYIGSQNASRGRORRHHPNVSQCGGCATCSDYNGCLSKPR 60
Db 1 MHLRLISWLPILNFMFYIGSQNASRGRORRHHPNVSQCGGCATCSDYNGCLSKPR 60
Qy 61 LPFVLERIGMKQIGVCLSSCPGSGYGYTRYPDINKTKCKVDCDTCFKNKFCCKSGFYL 120
Db 61 LPFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKTKCKADCDTCFKNKFCCKSGFYL 120
Qy 121 HLKCLDSCPEGLEANNHTECVSIHVCSEASWSPSCMKKGTCTGFKGTETVRDIL 180
Db 121 HLKCLDNCPEGLEANNHTECVSIHVCSEWNPSPCTKKGKTCGFKGTETVRREII 180
Qy 181 QHPSAKGKGNLCPPTSETRTCTIVQRKCKSGRGGKGRERKRLKLNKBERKETSSSDSK 240
Db 181 QHPSA--KGNLCPPTNETRKTIVQRKCKQKGRGGKGRERKRLKPNKGSKE--AIPDSK 236
Qy 241 GLESSIETPDQENKRRQOQKRRARDKQKSVSVSTVH 279
Db 237 SLESSKEIPEQRENK--QQKKRKVDK-QKSVSVSTVH 272

RESULT 4

US-10-125-852-23

; Sequence 23, Application US/10125852

; Publication No. US20030032034A1

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH FACTOR-1

; FILE REFERENCE: 28110/125,852

; CURRENT APPLICATION NUMBER: US/10/125,852

; CURRENT FILING DATE: 2002-08-20

; PRIOR APPLICATION NUMBER: US 60/316,368

; PRIOR FILING DATE: 2001-08-30

; PRIOR APPLICATION NUMBER: US 09/799,451

; PRIOR FILING DATE: 2001-03-05

; NUMBER OF SEQ ID NOS: 25

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 23

; LENGTH: 272

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-125-852-23

Query Match 85.7%; Score 1315.5; DB 14; Length 272;
Best Local Similarity 87.1%; Pred. No. 3.8e-97;
Matches 243; Conservative 11; Mismatches 18; Indels 7; Gaps 4;

Qy 1 MHLRLISCFPIILNFMFMEYIGSQNASRRQRMRHNPVNSQCGGCATCSYNGCLSKPR 60
Db 1 MHLRLISWLFILNFMFMEYIGSQNASRRQRMRHNPVNSQCGGCATCSYNGCLSKPR 60

Qy 61 LFFVLERIGMKQIGVCLSSCPSPGYGTRYPDINKTKKVDCTCFNKNFCTKCKSGFYL 120
Db 61 LFFALERIGMKQIGVCLSSCPSPGYGTRYPDINKTKKADCTCFNKNFCTKCKSGFYL 120

Qy 121 HLKCLDSCPEGLNANNHMECVSIHVCEASEWSPSCMKKGTCTGPKRGTTETVRDIL 180
Db 121 HLKCLDNCPEGLNANNHMECVSIHVCESEWNPSPCTKGTCTGPKRGTTETVRREII 180

Qy 181 QHPSAKGKGNLCPTSETRTCTIVQRKCKSGRGGKRRKRRKRLNKKERKETSSESSDSK 240
Db 181 QHPSA--KGNLCPTNETKCTVQRKCKSGRGGKRRKRRKRPKNGESKE--AIPDSK 236

Qy 241 GLESSIETPDQENKRRQOQKRRARDKQKSVSVSTVH 279
Db 237 SLESSKEIPEQRENK--QOQKRRKVQDK-QKSVSVSTVH 272

RESULT 5
US-09-894-912A-13
; Sequence 13, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; CURRENT FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-894-912A-13

Query Match 85.4%; Score 1310.5; DB 10; Length 273;
Best Local Similarity 87.1%; Pred. No. 9.6e-97;
Matches 242; Conservative 11; Mismatches 18; Indels 7; Gaps 4;

Qy 2 HRLRLISCFPIILNFMFMEYIGSQNASRRQRMRHNPVNSQCGGCATCSYNGCLSKPR 61
Db 3 HRLRLISWLFILNFMFMEYIGSQNASRRQRMRHNPVNSQCGGCATCSYNGCLSKPR 62

Qy 62 PFVLERIGMKQIGVCLSSCPSPGYGTRYPDINKTKKVDCTCFNKNFCTKCKSGFYL 121
Db 63 PFALERIGMKQIGVCLSSCPSPGYGTRYPDINKTKKADCTCFNKNFCTKCKSGFYL 122

Qy 122 LGKCLDSCPEGLNANNHMECVSIHVCEASEWSPSCMKKGTCTGPKRGTTETVRDILQ 181
Db 123 LGKCLDNCPEGLNANNHMECVSIHVCESEWNPSPCTKGTCTGPKRGTTETVRREIIQ 182

Qy 182 HPSAKGKGNLCPTSETRTCTIVQRKCKSGRGGKRRKRRKRLNKKERKETSSESSDSK 241
Db 181 QHPSA--KGNLCPTNETKCTVQRKCKSGRGGKRRKRRKRPKNGESKE--AIPDSK 236

Db 183 HPSA--KGNLCPTNETKCTVQRKCKSGRGGKRRKRRKRPKNGESKE--AIPDSK 238

Qy 242 LESSIETPDQENKRRQOQKRRARDKQKSVSVSTVH 279
Db 239 SLESSKEIPEQRENK--QOQKRRKVQDK-QKSVSVSTVH 273

RESULT 6
US-09-745-763-166
; Sequence 166, Application US/09745763
; Patent No. US20020065394A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; McCoy, John M.
; LaVallie, Edward R.
; Collins-Kacie, Lisa A.
; Evans, Cheryl
; Merberg, David
; Treacy, Maurice
; Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 219
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/745,763
; FILING DATE: 18-Jun-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 166:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 292 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 166:
US-09-745-763-166

Query Match 84.2%; Score 1293; DB 9; Length 292;
Best Local Similarity 85.1%; Pred. No. 2.6e-95;
Matches 235; Conservative 14; Mismatches 21; Indels 6; Gaps 3;

Qy 1 MHLRLISCFPIILNFMFMEYIGSQNASRRQRMRHNPVNSQCGGCATCSYNGCLSKPR 60
Db 1 MHLRLISWLFILNFMFMEYIGSQNASRRQRMRHNPVNSQCGGCATCSYNGCLSKPR 60

Qy 61 LFFVLERIGMKQIGVCLSSCPSPGYGTRYPDINKTKKVDCTCFNKNFCTKCKSGFYL 120
Db 61 LFFALERIGMKQIGVCLSSCPSPGYGTRYPDINKTKKADCTCFNKNFCTKCKSGFYL 120

Qy 121 HLKCLDSCPEGLNANNHMECVSIHVCEASEWSPSCMKKGTCTGPKRGTTETVRDIL 180
Db 121 HLKCLDNCPEGLNANNHMECVSIHVCESEWNPSPCTKGTCTGPKRGTTETVRREII 180

Qy 181 QHPSAKGKGNLCPTSETRTCTIVQRKCKSGRGGKRRKRRKRLNKKERKETSSESSDSK 240
Db 181 QHPSA--KGNLCPTNETKCTVQRKCKSGRGGKRRKRRKRPKNGESKE--AIPDSK 238

Db 181 QHPSA--KGNLCPPTNETRKTCTVQRKKCKQGERGKGRKRKPKNKGBSKB--AIPDSK 236
Qy 241 GLESSIETPDQENKRGQOQKRRARDKQKSVSVS 276
Db 237 SLESSKEIPEQRENK--QOQKRRKQVQDKQKSGIEVT 270

RESULT 7

US-09-894-912A-48
; Sequence 48, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; TITLE OF INVENTION: FACTOR-LIKE POLYPEPTIDES AND POLYNUCLEOTIDES
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; PRIOR FILING DATE: 2002-05-10
; PRIOR FILING DATE: 2001-04-05
; PRIOR FILING DATE: 2001-02-05
; PRIOR FILING DATE: 2001-06-28
; PRIOR FILING DATE: 2000-06-28
; PRIOR FILING DATE: 2001-01-09
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 48
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-894-912A-48

Query Match 84.2%; Score 1293; DB 10; Length 292;
Best Local Similarity 85.1%; Pred. No. 2.6e-95;
Matches 235; Conservative 14; Mismatches 21; Indels 6; Gaps 3;
Qy 1 MHLRLISCFPIILNFMFYIGSQNASRRGRORRHPNVSVSQCGGCATCSYNGCLSCPKR 60
Db 1 MHLRLISWLFILNFMFYIGSQNASRRGRORRHPNVSVSQCGGCATCSYNGCLSCPKR 60
Qy 61 LFFVLERIGMKQIGVCLSSCPSGYGTRYDPDINKTKCKVDCDTCFNNKPFCTCKSGFYL 120
Db 61 LFFALERIGMKQIGVCLSSCPSGYGTRYDPDINKTKCKADCDTCFNNKPFCTCKSGFYL 120
Qy 121 HLKCLDSCPEGLEANNHTMECVSIHVCHASSEWSPWPCMKKGTCTGFKGTETRVRDIL 180
Db 121 HLKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPWSPCTKGTCTGFKGTETRVREII 180
Qy 181 QHPSAKGNLCPPTSTRTCTVQRKKCKQGERGKGRKRKPKNKGBSKB--AIPDSK 240
Db 181 QHPSA--KGNLCPPTNETRKTCTVQRKKCKQGERGKGRKRKPKNKGBSKB--AIPDSK 236
Qy 241 GLESSIETPDQENKRGQOQKRRARDKQKSVSVS 276
Db 237 SLESSKEIPEQRENK--QOQKRRKQVQDKQKSGIEVT 270

RESULT 8

US-10-125-852-25
; Sequence 25, Application US/10125852
; Publication No. US20030032034A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH FACTOR-L
; TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES
; FILE REFERENCE: FVS-43A
; CURRENT APPLICATION NUMBER: US/10/125,852
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: US 60/316,368

; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: US 09/799,451
; BEST FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-125-852-25

Query Match 84.2%; Score 1293; DB 14; Length 292;
Best Local Similarity 85.1%; Pred. No. 2.6e-95;
Matches 235; Conservative 14; Mismatches 21; Indels 6; Gaps 3;
Qy 1 MHLRLISCFPIILNFMFYIGSQNASRRGRORRHPNVSVSQCGGCATCSYNGCLSCPKR 60
Db 1 MHLRLISWLFILNFMFYIGSQNASRRGRORRHPNVSVSQCGGCATCSYNGCLSCPKR 60
Qy 61 LFFVLERIGMKQIGVCLSSCPSGYGTRYDPDINKTKCKVDCDTCFNNKPFCTCKSGFYL 120
Db 61 LFFALERIGMKQIGVCLSSCPSGYGTRYDPDINKTKCKADCDTCFNNKPFCTCKSGFYL 120
Qy 121 HLKCLDSCPEGLEANNHTMECVSIHVCHASSEWSPWPCMKKGTCTGFKGTETRVRDIL 180
Db 121 HLKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPWSPCTKGTCTGFKGTETRVREII 180
Qy 181 QHPSAKGNLCPPTSTRTCTVQRKKCKQGERGKGRKRKPKNKGBSKB--AIPDSK 240
Db 181 QHPSA--KGNLCPPTNETRKTCTVQRKKCKQGERGKGRKRKPKNKGBSKB--AIPDSK 236
Qy 241 GLESSIETPDQENKRGQOQKRRARDKQKSVSVS 276
Db 237 SLESSKEIPEQRENK--QOQKRRKQVQDKQKSGIEVT 270

RESULT 9

US-09-894-912A-26
; Sequence 26, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; TITLE OF INVENTION: FACTOR-LIKE POLYPEPTIDES AND POLYNUCLEOTIDES
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; CURRENT FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 26
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-894-912A-26

Query Match 83.7%; Score 1285; DB 10; Length 265;
Best Local Similarity 86.7%; Pred. No. 1e-94;
Matches 234; Conservative 12; Mismatches 18; Indels 6; Gaps 3;
Qy 1 MHLRLISCFPIILNFMFYIGSQNASRRGRORRHPNVSVSQCGGCATCSYNGCLSCPKR 60
Db 1 MHLRLISWLFILNFMFYIGSQNASRRGRORRHPNVSVSQCGGCATCSYNGCLSCPKR 60

Qy 61 LFFVLERIGMKQIGVCLSSCPGSGYGTGTRYPDINKTKCKVDCTCFNKNFCTKCKSGPYL 120
Db 61 LFFALERIGMKQIGVCLSSCPGSGYGTGTRYPDINKTKCKADCTCFNKNFCTKCKSGPYL 120
Qy 121 HLGKCLDSCPEGLBANNHTMVCVSVHCEASWSPCMKGGKTCGPKRGTTETVRDIL 180
Db 121 HLGKCLDNCPEGLBANNHTMVCVSVHCEVSEWNPSPCTKGGKTCGPKRGTTETVRBII 180
Qy 181 QHPSAKGKGLCPPTSETTCTVQRKCKSGRGGKGRKRRKKNLKNKERTSSSSDSK 240
Db 181 QHPSA--KGNLCPTTNETRKTCTVQRKCKGKGRKRRKKNKNGESKE--AIPDSK 236
Qy 241 GLESSIETPDQENKERQKQKRRARDKQ 270
Db 237 SLESSKEIPEQRENK--QKQKRVQDKOK 264

RESULT 10

US-09-894-912A-16
; Sequence 16, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; CURRENT FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 16
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-894-912A-16

Query Match 79.6%; Score 1222.5; DB 10; Length 251;
Best Local Similarity 86.8%; Pred. No. 9.3e-90;
Matches 224; Conservative 11; Mismatches 16; Indels 7; Gaps 4;
Qy 22 QNAGRGRQRMRHNPVNSGCGGCATCSYNGCLSCPKRLFPVLRIGMKQIGVCLSSCP 81
Db 1 QNAGRGRQRMRHNPVNSGCGGCATCSYNGCLSCPKRLFPALERIGMKQIGVCLSSCP 60
Qy 82 SGYGTGTRYPDINKTKCKVDCTCFNKNFCTKCKSGPYLHLGKCLDSCPEGLBANNHTME 141
Db 61 SGYGTGTRYPDINKTKCKADCTCFNKNFCTKCKSGPYLHLGKCLDNCPEGLBANNHTME 120
Qy 142 CVSVHCEASWSPCMKGGKTCGPKRGTTETVRDILQHPSAKGKGLCPPTSETTCTC 201
Db 121 CVSVHCEVSEWNPSPCTKGGKTCGPKRGTTETVRBIIQHPSA--KGNLCPTTNETRKC 178
Qy 202 IVQRKCKSGRGGKGRKRRKKNLKNKERTSSSSDSKGLSSLETTPDQENKERQKQ 261
Db 179 TVQRKCKQGRGGKGRKRRKKNKNGESKE--AIPDSKSKLESSKEIPEQRENK--QKQK 234
Qy 262 KRRARDKQKSVSVSTVH 279
Db 235 KRVQDK--QKSVSVSVSTVH 251

RESULT 11

US-10-185-770-4

; Sequence 4, Application US/10185770
; Publication No. US20030022217A1
; GENERAL INFORMATION:
; APPLICANT: CECCARDI, Toni et al.
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; FILE REFERENCE: CL0001247
; CURRENT APPLICATION NUMBER: US/10/185,770
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/301,852
; PRIOR FILING DATE: 2001-07-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-185-770-4
Query Match 77.3%; Score 1187; DB 14; Length 225;
Best Local Similarity 92.5%; Pred. No. 5.7e-87;
Matches 210; Conservative 5; Mismatches 10; Indels 2; Gaps 1;
Qy 1 MHLRLISCFPIILNFMVEYIGSONASRGRRQRMRHNPVNSGCGGCATCSYNGCLSCPKR 60
Db 1 MHLRLISWLFPIILNFMVEYIGSONASRGRRQRMRHNPVNSGCGGCATCSYNGCLSCPKR 60
Qy 61 LFFVLERIGMKQIGVCLSSCPGSGYGTGTRYPDINKTKCKVDCTCFNKNFCTKCKSGPYL 120
Db 61 LFFALERIGMKQIGVCLSSCPGSGYGTGTRYPDINKTKCKADCTCFNKNFCTKCKSGPYL 120
Qy 121 HLGKCLDSCPEGLBANNHTMVCVSVHCEASWSPCMKGGKTCGPKRGTTETVRDIL 180
Db 121 HLGKCLDNCPEGLBANNHTMVCVSVHCEVSEWNPSPCTKGGKTCGPKRGTTETVRBII 180
Qy 181 QHPSAKGKGLCPPTSETTCTVQRKCKSGRGGKGRKRRKKNLKN 227
Db 181 QHPSA--KGNLCPTTNETRKTCTVQRKCKGKGRKRRKKNKPNK 225

RESULT 12

US-10-192-1782
; Sequence 1782, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1782
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-087-192-1782

Query Match 73.9%; Score 1135; DB 12; Length 239;
Best Local Similarity 90.1%; Pred. No. 8.6e-83;
Matches 201; Conservative 6; Mismatches 14; Indels 2; Gaps 1;
Qy 1 MHLRLISCFPIILNFMVEYIGSONASRGRRQRMRHNPVNSGCGGCATCSYNGCLSCPKR 60
Db 18 MHLRLISWLFPIILNFMVEYIGSONASRGRRQRMRHNPVNSGCGGCATCSYNGCLSCPKR 77

```

; NUMBER OF SEQ ID NOS: 298
; SOFTWARE: PatentIn 2.1
; SEQ ID NO 172
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-886-172

```

Query Match	71.7%;	Score 1100;	DB IS;	Length 4359;
Best Local Similarity	90.4%;	Pred. No. 5.4e-80;		
Matches 197;	Conservative	5; Mismatches 10;	Indels 6;	Gaps 3;
QY	1	MHLRLISCFPIILNFMEYIGSQNASRGRQRORRHPNPVNSQCGGCATCSDYNGCLSKCPKPR	60	
DB	1	MHLRLISLWFIILNFMEYIGSQNASRGRQRORRHPNPVNSQCGGCATCSDYNGCLSKCPKPR	60	
QY	61	LPFVLERIGWKQIGVCLSSCPSGYGTRYDPDINKCT-KC---KVDCTCFNKNFCTCKS	116	
DB	61	LPFALERIGWKQIGVCLSSCPSGYGTRYDPDINKCTSKCPHEKADCTCFNKNFCTCKS	120	
QY	117	GFYHLHGLKCLDSCPEGLRANHTWECVSIYHCEASEWSPMSCKGKTCGPKRGGTETRV	176	
DB	121	GFYHLHGLKCLDNCPEGLRANHTWECVSSVHCESEWNPSCTYKKGKTCGPKRGGTETRV	180	
QY	177	RDILQHPSAKGNLCPPTSETRTCTIVQRKKCSKRG	214	
DB	181	REIIQHPSA--KGNLCPPTNTRKCTVQRKKCQKRG	216	

RESULT 14
US-10-094-886-176
Application US/10094886

US-10-094-886-1176
 * Sequence 176, Application US/10094886
 * Publication No. US20040002120A1
 * GENERAL INFORMATION:
 * APPLICANT: Kektuda, Ramesh
 * APPLICANT: Tchernev, Velizar T.
 * APPLICANT: Liu, Xiaochong
 * APPLICANT: Spytchak, Kimberly A.
 * APPLICANT: Patturajan, Meera
 * APPLICANT: Burgess, Catherine
 * APPLICANT: Varnet, Corine A.
 * APPLICANT: Li, Li
 * APPLICANT: Gorman, Linda
 * APPLICANT: Malyankar, Uriel M.
 * APPLICANT: Boldog, Ferenc
 * APPLICANT: Guo, Xiaojia
 * APPLICANT: Shenoy, Suresh
 * APPLICANT: Padigar, Muralidhara
 * APPLICANT: Padigar, Raymond J., Jr.

APPLICANT: LAROCHE, William
APPLICANT: Zhong, Mei

FILE REFERENCE: 21402-290 B
 INVESTIGATION NUMBER: US/10/094.886

CURRENT APPLICATION NUMBER: US/107
FILING DATE: 2002-03-07

CURRENT FILING DATE: 2002-03-07
REGISTRATION NUMBER: 601274

PRIOR APPLICATION NUMBER: 60/274

PRIOR FILING DATE: 2001-03-08
PRIORITY DATE: 2001-03-08

PRIOR APPLICATION NUMBER: 80/313-
PUBLISHING DATE: 2001-08-17

PRIOR FILING DATE: 2001-08-17
PRIORITY APPLICATION NUMBER: 60/288

PRIOR APPLICATION NUMBER: 00/233
PRIOR FILING DATE: 2001-05-02

PRIOR FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: 60/318

PRIOR APPLICATION NUMBER: 00/513
PRIOR FILING DATE: 2001-09-10

PRIOR FILING DATE: 2002-01-01
PRIOR APPLICATION NUMBER: 60/274

PRIOR FILING DATE: 2001-03-08

PRIOR FILING DATE: 60/314
PRIOR APPLICATION NUMBER: 60/314

PRIOR FILING DATE: 2001-08-21

PRIOR APPLICATION NUMBER: 60/274

PRIOR FILING DATE: 2001-03-08

PRIOR APPLICATION NUMBER: 60/274

PRIOR FILING DATE: 2001-03-09

PRIOR APPLICATION NUMBER: 60/296607

PRIOR FILING DATE: 2001-06-07

PRIOR APPLICATION NUMBER: 60/315

PRIOR FILING DATE: 2001-08-21

Remaining Prior Application data

```
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/318,510
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/314,018
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/296,693
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/313,626
; PRIOR FILING DATE: 2001-08-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 298
; SOFTWARE: PatentIn 2.1
; SEQ ID NO 176
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-886-176

Query Match      68.1%; Score 1045; DB 15; Length 195;
Best Local Similarity 92.8%; Pred. No. 1.1e-75;
Matches 181; Conservative 5; Mismatches 7; Indels 2; Gaps 1;

QY 20 GSONASRRQRMRHNPVSGCGGCATCSDYNGCLSCPKRLFPVLERIGMKQIGVCLSS 79
Db 1 GSONASRRQRMRHNPVSGCGGCATCSDYNGCLSCPKRLFPVLERIGMKQIGVCLSS 60

QY 80 CPSGYGTYPDINKCTCKVDCDTCFNKNFCTCKSGFVHLGKCLDSCPEGLNANH 139
Db 61 CPSGYGTYPDINKCTCKVDCDTCFNKNFCTCKSGFVHLGKCLDSCPEGLNANH 120

QY 140 MECVSIHVCEASEWSPWCMKKGKTCGPKRGTTETVRDILQHPSAKGNLCPTSETR 199
Db 121 MECVSIHVCEASEWSPWCMKKGKTCGPKRGTTETVRDILQHPSAKGNLCPTSETR 178

QY 200 TCIVQRKKCKSGERG 214
Db 179 KCTVQRKKCKSGERG 193

RESULT 15
US-10-094-886-174
; Sequence 174, Application US/10094886
; Publication No. US20040002120A1
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Patturajan, Meera
; APPLICANT: Burgess, Catherine
; APPLICANT: Vernet, Corine A.
; APPLICANT: Li, Li
; APPLICANT: Gorman, Linda
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Boldog, Ferenc
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shenoy, Suresh
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Taupier, Raymond J., Jr.
; APPLICANT: Miller, Charles
; APPLICANT: Casman, Stacie
; APPLICANT: Pena, Carol
; APPLICANT: Gangolli, Esha
; APPLICANT: Gusev, Vladimir
; APPLICANT: Smithson, Glennnda
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Gerlach, Valerie
```

```
; APPLICANT: Pochart, Pascal
; APPLICANT: Fernandes, Elma
; APPLICANT: Shinkets, Richard
; APPLICANT: Rastelli, Luca
; APPLICANT: Spaderna, Steven
; APPLICANT: Larochele, William
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-290 B US/10/094,886
; CURRENT APPLICATION NUMBER: 2002-03-07
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/274,322
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/313,182
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/288,052
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/318,510
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/314,018
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/296,693
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/313,626
; PRIOR FILING DATE: 2001-08-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 298
; SOFTWARE: PatentIn 2.1
; SEQ ID NO 174
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-886-174
```

```
Query Match      67.8%; Score 1041; DB 15; Length 195;
Best Local Similarity 92.3%; Pred. No. 2.2e-75;
Matches 180; Conservative 6; Mismatches 7; Indels 2; Gaps 1;

QY 20 GSONASRRQRMRHNPVSGCGGCATCSDYNGCLSCPKRLFPVLERIGMKQIGVCLSS 79
Db 1 GSONASRRQRMRHNPVSGCGGCATCSDYNGCLSCPKRLFPVLERIGMKQIGVCLSS 60

QY 80 CPSGYGTYPDINKCTCKVDCDTCFNKNFCTCKSGFVHLGKCLDSCPEGLNANH 139
Db 61 CPSGYGTYPDINKCTCKVDCDTCFNKNFCTCKSGFVHLGKCLDSCPEGLNANH 120

QY 140 MECVSIHVCEASEWSPWCMKKGKTCGPKRGTTETVRDILQHPSAKGNLCPTSETR 199
Db 121 MECVSIHVCEASEWSPWCMKKGKTCGPKRGTTETVRDILQHPSAKGNLCPTSETR 178

QY 200 TCIVQRKKCKSGERG 214
Db 179 KCTVQRKKCKSGERG 193
```

Search completed: June 29, 2004, 17:15:57
Job time : 37.5837 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: June 29, 2004, 16:56:05 ; Search time 14.2209 Seconds
(without alignments)
1887.186 Million cell updates/sec

Title: US-09-894-912A-32
Perfect score: 1535
Sequence: 1 MLRLISCPFIILFMFYIG.....QQRRARDKQKSVSVTVH 279
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	187	12.2	1548	2 S34583	serine proteinase
2	185.5	12.1	1299	2 T43251	furin (EC 3.4.21.7)
3	179	11.7	969	1 A39490	subtilisin-like pr
4	174.5	11.4	962	2 JC5571	subtilisin-like pr
5	174.5	11.4	975	2 JC5570	subtilisin-like pr
6	174	11.3	1680	2 A43434	furin (EC 3.4.21.7)
7	171	11.1	932	2 I52527	PACB4A - mouse (fr
8	170.5	11.1	942	2 D87803	protein bli-4D [im
9	164	10.7	915	2 B48225	probable protease
10	163.5	10.7	915	1 A48225	subtilisin-like pr
11	162	10.6	570	2 T37314	probable kexin (EC
12	161	10.5	937	2 I53282	gene PACB4 protein
13	154	10.0	899	2 G02428	subtilisin-like pr
14	154	10.0	915	2 JC6148	subtilisin-like pr
15	149	9.7	1371	2 A33837	insulin-like growth
16	145.5	9.5	803	2 A47723	P-spondin precursor
17	144.5	9.4	2471	2 A49128	cell fate determin
18	142.5	9.3	807	2 A38152	P-spondin - rat
19	140.5	9.2	1203	2 A49175	Motch B protein -
20	140	9.1	378	2 B59180	Wat inhibitory fac
21	139.5	9.1	1367	1 IGHUR1	insulin-like growth
22	139	9.1	2524	2 A35844	Xotch protein - Af
23	138.5	9.0	379	2 A59180	Wat inhibitory fac
24	137	8.9	440	2 T24232	hypotheical prote
25	137	8.9	1620	2 T27283	hypotheical prote
26	136.5	8.9	1369	2 S70713	protein-tyrosine k
27	133	8.7	540	2 B47417	insulin receptor-r
28	131	8.5	1268	2 B65502	insulin receptor-r
29	131	8.5	2101	2 S57245	insulin receptor (

30 131 8.5 2148 1 A56081 insulin receptor -
31 130.5 8.5 1382 1 INHUR insulin receptor p
32 130 8.5 2531 2 S18188 notch protein homo
33 130 8.5 2531 2 A46019 notch-1 protein -
34 129.5 8.4 1372 2 A34157 insulin receptor p
35 129.5 8.4 1383 2 A36080 insulin receptor p
36 129 8.4 2437 2 S42612 transmembrane prot
37 129 8.4 2555 2 A40043 notch protein homo
38 128.5 8.4 1111 2 T26972 hypothetical prote
39 128 8.3 461 2 JC4302 tumor necrosis fac
40 126.5 8.2 837 2 S43656 furin (EC 3.4.21.7
41 126 8.2 677 2 C42125 trophozoite cystei
42 125.5 8.2 1274 2 T42017 trophozoite surfac
43 124.5 8.1 667 2 A48579 trophozoite surfac
44 124 8.1 1300 2 A36502 insulin receptor-r
45 123.5 8.0 375 2 S26059 probable transform

ALIGNMENTS

RESULT 1

S34583
serine proteinase (EC 3.4.21.-) PC6B - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 05-Nov-1999
C:Accession: S34583
R:Nakagawa, T.; Murakami, K.; Nakayama, K.
FEBS Lett. 327, 165-171, 1993
A:Title: Identification of an isoform with an extremely large Cys-rich region of PC6, a
A:Reference number: S34583; MUID:93327934; PMID:8335106
A:Accession: S34583
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1548 <NAK>
A:Cross-references: GB:D17583; NID:G407344; PIDN:BAA04507.1; PID:dl005033; PID:G440374
C:Keywords: hydrolase; serine proteinase

Query Match 12.2%; Score 187; DB 2; Length 1548;
Best Local Similarity 27.8%; Pred. NO. 0.0001;
Matches 52; Conservative 21; Mismatches 72; Indels 42; Gaps 11;

QY 38 SQCGGSGCATCSYNG--CLSCRPRLFVFLERIGMKQIGVCLSSCPGSGYGYTRYDPI--N 93
DB 880 NQCHSSCKTC---NGSLCASCPTGMYLWQ-----ACVPSCPQG---TWPSVTSG 924
QY 94 KCTKCKYVDCPTFNKNPCTCKS---GFYLHLGKLDSCPGLEANNHTNHCVSIVHCE 149
DB 925 SCEKSEDCVSCGADICQQCLSQPDNTLLHRCRCYHSCPEGFYAKGVCE-----HCS 979
QY 150 ASWSPWSPCKKGTCTGFRGTETRTVRDIL-----QHPSAKGNLCPTTSETRTCTIV 203
DB 980 ----SPCKTCEGNATSCNCEGDFVLDHGVCKWCTCPKHAIVEGVCKHCP---ERCQDCIH 1033
QY 204 QR--KCC 208
DB 1034 EKTCKEC 1040

RESULT 2

T43251
furin (EC 3.4.21.75) - fall armyworm
N:Alternate names: paired basic amino acid cleaving enzyme; proprotein convertase; serin
C:Species: Spodoptera frugiperda (fall armyworm)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
C:Accession: T43251
R:Cieplik, M.; Klenk, H.
submitted to the EMBL Data Library, January 1996
A:Description: Cloning and functional characterization of PURIN from Spodoptera frugiper
A:Reference number: 222368
A:Accession: T43251
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA

A;Cross-references: GDB:131390; OMIM:167405

A;Map position: 15q26-15q26

C;Superfamily: subtilisin-like proteinase PACE4; subtilisin homology

C;Keywords: alternative splicing; glycoprotein; hydrolase; serine proteinase

F;1-62/Domain: signal sequence #status predicted <SIG>

F;63-149/Domain: propeptide #status predicted <PRO>

F;196-434/Domain: subtilisin homology <SBT>

F;952-968/Domain: hydrophobic cluster #status predicted <HCL>

F;205-246,347,420/Active site: Asp, His, Ser #status predicted

F;259/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 11.4%; Score 174.5; DB 2; Length 975;

Best Local Similarity 23.9%; Pred. No. 0.00045;

Matches 53; Conservative 25; Mismatches 81; Indels 63; Gaps 10;

QY 18 YIGSQNASRRRRMHPNVSQCGGCATCSD--YNGCLSKPRLFFVLERIGMKQIGV 75

|||||

Db 737 YFGDTAARRCR-----CHKGCTCSSRAATQCLSCR-RGPF-----HHQEMNT 779

|||||

QY 76 CLSSCPGGYGTTRYPDINKTKKVDCTCFNK-NFTCKSGFYHLGKCLDSCPEGLE 134

|||||

Db 780 CVTLCPAGFYADE--SQKNCLKHPCKKCVDRPEKCTVCKEGFSLARGSCIPDCEPGTY 837

|||||

QY 135 ANNHTMEC-----VSIHVCRAS-----EWS-----PWSP 158

|||||

Db 838 FDSRLRGCHHTGTCVGPGRBECIHCAKNFPHDWKVCVPACGSGFYPERMPGLPHKV 897

|||||

QY 159 CMKKGKTCGPKRGTRTV-----RDILQHPSPAKGKGNLCPPPT 195

|||||

Db 898 CRRYPPGGBERQATVSSKGVPGQSLSSASSPGAGGMLHHT 939

|||||

RESULT 6

A43434

furin (BC 3.4.21.75) 2 - fruit fly (Drosophila melanogaster)

C;Species: Drosophila melanogaster

C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 31-Mar-2000

C;Accession: A43434

R;Roebroek, A.J.; Creemers, J.W.; Pauli, I.G.; Kurzik-Dumke, U.; Rentrop, M.; Gateff, E.

J. Biol. Chem. 267, 17208-17215, 1992

A;Title: Cloning and functional expression of Dfurin2, a subtilisin-like proprotein proc

A;Reference number: A43434; MUID:92381036; PMID:1512259

A;Accession: A43434

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1680 <ROE>

A;Cross-references: GB:M94375; NID:g157461; PID:g157462

A;Note: sequence extracted from NCBI backbone (NCBIN:111933, NCBIIP:111934)

C;Genetics:

A;Gene: FlyBase:Fur2

A;Cross-references: FlyBase:FBgn0004598

C;Superfamily: subtilisin homology

C;Keywords: hydrolase; serine proteinase; transmembrane protein

F;409-652/Domain: subtilisin homology <SBT>

F;418,457,638/Active site: Asp, His, Ser #status predicted

Query Match 11.3%; Score 174; DB 2; Length 1680;

Best Local Similarity 24.5%; Pred. No. 0.00073;

Matches 48; Conservative 15; Mismatches 61; Indels 72; Gaps 6;

QY 23 NASRRRRMHPN--VSQCGGCATCSDY--NGCLSKPRLFFVLERIGMKQIGVCL 77

|||||

Db 995 NTCVSRCPRRFPNQVGICWPCHDTCACAGPDSCLTCAPAHLHVID-----LAVCL 1048

|||||

QY 78 SSCPSGY-----GTRYPDI 92

|||||

Db 1049 QFCDGYFENSRNRTCTVCEPNCASQDHPYCTSCDHLVMHEKCYSAACPLDITYETD 1108

|||||

QY 93 NKCTCKKVDCTCF--NKNFTCKSGFYHLGKCLDSCPEGLEANNHTMECVSIHVCR 150

|||||

Db 1109 NKCAFCHSTCATCGPTDQDCLTCRSSRYAWQNKCLISCDPGFYADKKRLSCM----- 1161

QY 151 SEWSPWSPCKMKKTC 166

Db 1162 -----PCQEGCKYC 1170

|||||

RESULT 7

I52527

PACB4A - mouse (fragment)

C;Species: Mus musculus (house mouse)

C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 16-Jul-1999

C;Accession: I52527

R;Hosaka, M.; Murakami, K.; Nakayama, K.

Biomed. Res. 15, 383-390, 1994

A;Title: PACEA is a ubiquitous endoprotease that has similar but not identical substrat

A;Accession: I52527

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-932 <RES>

A;Cross-references: GB:D50060; NID:g769700; PID:BAA08777.1; PID:g769701

C;Superfamily: subtilisin-like proteinase PACE4; subtilisin homology

F;172-410/Domain: subtilisin homology <SBT>

Query Match 11.1%; Score 171; DB 2; Length 932;

Best Local Similarity 22.9%; Pred. No. 0.00073;

Matches 55; Conservative 25; Mismatches 88; Indels 72; Gaps 12;

QY 18 YIGSQNASRRRRMHPNVSQCGGCATCSDY--GCLSKPRLFFVLERIGMKQIGV 75

|||||

Db 700 YFGDAAARRCR-----CHKGCTCTGRSPAQCLSCR-RGPF-----HHQEMNT 742

|||||

QY 76 CLSSCPGGYGTTRYPDINKTKKVDCTCFNK-NFTCKSGFYHLGKCLDSCPEGLE 134

|||||

Db 743 CVTLCPAGLYADESQRL--CLRCHPSCKQCVDEPEKCTVCKEGFSLARGSCIPDCEPGTY 800

|||||

QY 135 ANNHTMEC-----VSIHVCRAS-----EWS-----PWSP 158

|||||

Db 801 FDSRLVKGCHHTCTCTCVGPSRECIHCAKSPHFQDWKVCVPACGSGFYPERMPGLPHKV 860

|||||

QY 159 CMKKGKTC-----GPKRG-----TTRVDRILQHPSPAKGKGNLCPPPTSETTRTCIVORK 206

|||||

Db 861 CRRCEENCLSCGSSRNCRCKAGFTQLGTSCTIHTNHTCSNADETFCWVKSNNRLC--BRK 918

RESULT 8

D87803

protein bli-4D [imported] - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 21-Jun-2002

C;Accession: D87803

R;anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biologi

A;Reference number: A75000; MUID:99069613; PMID:9851916

A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_ele

A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and

A;Accession: D87803

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-942 <STO>

A;Cross-references: GB:chr_1; PIDN:AAB96754.1; PID:g2773243; GSPDB:GN000019

C;Genetics:

A;Gene: bli-4D

A;Map position: 1

C;Superfamily: kexin; subtilisin homology

Query Match 11.1%; Score 170.5; DB 2; Length 942;

Best Local Similarity 28.3%; Pred. No. 0.0008;

Matches 53; Conservative 18; Mismatches 75; Indels 41; Gaps 9;

QY 36 NVSGCGCGGCATCSDYNGCLSKPRLFFVLERIGMKQIGVCLSSCPSGYGTTRYPDINKC 95

|||||

Db 674 NCHDECGNGCTESSATSCFPACK-LLTQTLRNKGSGP-KCVQKCDITY-----YLDGDKC 727

QY 96 TRKVDCTCFNKNPCTCKSGPYL-----HL--GKCLDSCPGLBANNHTMBCVSI 146
DB 728 RWCSSCHCTTCAEVCECTCGSLLLIDVDNMPHYDGHCKVSCPPGLVADYBS-----NLV 783
QY 147 HCEASEWSPSPCKMKGKTCGPKRGTETRVDRILQHPSAKGNLCPPYSETRTC--IVQ 204
DB 784 QAK-----CIWRKDLGGDG-----YYINAVGKCDLDCSSCTCTAGCPMS 823
QY 205 RKCKSKG 211
DB 824 CEKCKSKG 830
RESULT 9
B48225
probable proprotein convertase (EC 3.4.21.-) 5 precursor - rat
N/Alternate names: PC5 precursor
C/Species: Rattus norvegicus (Norway rat)
C/Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 23-Feb-1997
C/Accession: B48225
R.;Jussan, J.; Vieau, D.; Hamelin, J.; Day, R.; Chretien, M.; Seidah, N.G.
Proc. Natl. Acad. Sci. U.S.A. 90, 6691-6695, 1993
A/Title: cDNA structure of the mouse and rat subtilisin/kexin-like PCS: a candidate pro-
A/Reference number: A48225; MUID:93342056; PMID:8341687
A/Accession: B48225
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-915 <LUS>
A/Cross-references: GB:114933
C/Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
C/Keywords: duplication; glycoprotein; hydrolase; integrin binding; serine proteinase
F:1-34/Domain: signal sequence #status predicted <SIG>
F:35-915/Product: probable proprotein convertase 5 #status predicted <PRO>
F:117-915/Product: probable proprotein convertase 5 #status experimental <MAT>
F:164-402/Domain: subtilisin homology <SST>
F:173,214,388/Active site: Asp, His, Ser #status predicted
Query Match 10.7%; Score 164; DB 2; Length 915;
Best Local Similarity 27.1%; Pred. No. 0.002;
Matches 48; Conservative 20; Mismatches 57; Indels 48; Gaps 9;
QY 35 PNVSQ--GCOQ-----GCATCSYNGCLSKPRLFFVLERIGMKQIGVCLSSCPGGYGT 89
DB 640 PCSEVGGDGPDPDCTCLHYHKLNNTR-----LCVSSCPGHP---H 682
QY 90 PDINKCTCKVDCDTCFNRNP--CTKCKSGFYHL--LGRKCLDSCPGLBANNHTMBCVSI 145
DB 683 ADKRCRCKAPNCESCFGSHADQCLSKYGFNLBETSCVACQCPGSGYQDIKKNIC--- 739
QY 146 VHCEASEWSPSPCKMKGKTC-GPKRGTETRVDRILQHPSAKGNLCPPYTSR 197
DB 740 -----GKCSNCKTCTGFHNCTCKGGLSLQ-----GSRCSVTCE 774
RESULT 10
A48225
subtilisin-like proprotein convertase (EC 3.4.21.-) PC5 precursor - mouse
N/Alternate names: kexin homolog; serine proteinase PC6
C/Species: Mus musculus (house mouse)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C/Accession: A48225; JX0248
R.;Jussan, J.; Vieau, D.; Hamelin, J.; Day, R.; Chretien, M.; Seidah, N.G.
Proc. Natl. Acad. Sci. U.S.A. 90, 6691-6695, 1993
A/Title: cDNA structure of the mouse and rat subtilisin/kexin-like PCS: a candidate pro-
A/Reference number: A48225; MUID:93342056; PMID:8341687
A/Accession: A48225
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-915 <LUS>
A/Cross-references: GB:114932; NID:9293327; PIDN:AAA74636.1; PID:9293328
J.Nakagawa, T.; Hosaka, M.; Torii, S.; Watanabe, T.; Murakami, K.; Nakayama, K.
J. Biochem. 113, 132-135, 1993
A/Title: Identification and functional expression of a new member of the mammalian Kex2-

A/Reference number: JX0248; MUID:93224489; PMID:8468318
A/Accession: JX0248
A/Molecule type: mRNA
A/Residues: 1-915 <NAK>
A/Cross-references: DDBJ:D12619; NID:9220565; PIDN:BA02143.1; PID:9220566
A/Note: the authors translated the codon GGC for residue 915 as Ala
C/Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
C/Keywords: duplication; glycoprotein; hydrolase; integrin binding; serine proteinase
F:1-34/Domain: signal sequence #status predicted <SIG>
F:35-116/Domain: propeptide #status predicted <PRO>
F:117-915/Product: proprotein convertase PC5 #status experimental <MAT>
F:164-402/Domain: subtilisin homology <SST>
F:173,214,388/Active site: Asp, His, Ser #status predicted
Query Match 10.7%; Score 163.5; DB 1; Length 915;
Best Local Similarity 29.1%; Pred. No. 0.0022;
Matches 44; Conservative 18; Mismatches 50; Indels 39; Gaps 8;
QY 20 GSQNASRGRQRMRHNPVSGCGCGCATCSYNG--CLSCPKRLFFVLERIGMKQIGVCL 77
DB 679 GHYHADK-KRCRCAPN-----CESCFGSHDQCLSKYGYFL-----NEETSSCV 723
QY 78 SSCPSGYGTRYPDINK--CTKCKVDCDTCFNRNFTCKSGPYHLGKCLDSCPGLB 135
DB 724 TQCPDG---SYEDIKKNVCGKCSNCKACIGFHNCTCKGGLSLQSGSRCSVTCDGQFF 779
QY 136 NNHTMBCVSIHCEASEWSPSPCKMKGKTC 166
DB 780 NGEH--DC-----QPCRFPCATC 794
RESULT 11
T37314
probable kexin (EC 3.4.21.61) - Caenorhabditis elegans (fragment)
N/Alternate names: bliesterase 4
C/Species: Caenorhabditis elegans
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
C/Accession: T37314
R.;Thacker, C.; Peters, K.; Srayko, M.; Rose, A.M.
Genes Dev. 9, 956-971, 1995
A/Title: The bli-4 locus of Caenorhabditis elegans encodes structurally distinct kex2/5
A/Reference number: Z21679; MUID:95293228; PMID:7774813
A/Accession: T37314
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-570 <THA>
A/Cross-references: EMBL:L29440; NID:9459702; PIDN:AAA98752.1; PID:9459703
C/Genetics:
A/Gene: bli-4
A/Map position: 1
C/Keywords: alternative splicing; hydrolase; serine proteinase
Query Match 10.6%; Score 162; DB 2; Length 570;
Best Local Similarity 27.1%; Pred. No. 0.0019;
Matches 54; Conservative 21; Mismatches 76; Indels 48; Gaps 10;
QY 31 RMHNPVSG-----COGGCATCSYNGCLSKPRLFFVLERIGMKQIGVCLSSCPG 83
DB 290 RSGFPDLTSGWKLSCDECGGCTESSATSCFAFK-HLTOTLANKGSGP-KCVQKDDT 347
QY 84 YGTRYPDINKCTCKVDCDTCFNRNFTCKSGFYL-----HL--GKCLDSCPGLB 134
DB 348 Y-----YLDGDKCMKCSHCHCTCTKAECVCTCPGSLLLIDVDNMPHYDGHCKVSCPPGLV 403
QY 135 ANNHTMBCVSIHCEASEWSPSPCKMKGKTCGPKRGTETRVDRILQHPSAKGNLCPP 194
DB 404 ADYES-----NLVQAK-----CIWRKDLGGDG-----YYINAVGKCDLDCS 439
QY 195 TSETRTC--IVQRCKCKSKG 211
DB 440 SCSTCTAPGPMSCCKCKSKG 458

RESULT 12

153282
Gene PACE4 protein - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 22-Jun-1999
C:Accession: I53282
R:Johnson, R.C.; Darlington, D.N.; Hand, T.A.; Bloomquist, B.T.; Mains, R.E.
Endocrinology 135, 1178-1185, 1994
A:Title: PACE4: a subtilisin-like endoprotease prevalent in the anterior pituitary and
A:Reference number: I53282; MUID:94349873; PMID:8070361
A:Accession: I53282
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-937 <RES>
A:Cross-references: GB:L31894; NID:g496221; PIDN:AAA61987.1; PID:g496222
C:Genetics:
A:Gene: PACE4
C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
P:177-415/Domain: subtilisin homology <SET>

Query Match 10.5%; Score 161; DB 2; Length 937;
Best Local Similarity 22.5%; Pred. No. 0.0032;
Matches 54; Conservative 24; Mismatches 90; Indels 72; Gaps 12;

QY 18 YISQNASRRQRBMHPNVISQCGCATCSQDYN--GCLSKPRLFFVLERIGMKQIGV 75

DB 705 YFGDTAARRCR-----CHKGCETCTGRSPQCLSCR-RGFI-----HHQSTNT 747

QY 76 CLSSCPGYYGTRYPDINKCTKCKVDCTCFNK-NFCTCKSGFYHLGKCLDSCPBLGLE 134

DB 748 CVTLCPAGLYADESQL--CLRHPSQCKVDPEKSTVCKEGLARGSCIPDCBPQTY 805

QY 135 ANNHTMEC-----VSIHVCEAS-----EWS-----PMSP 158

DB 806 FDSELRIGRGCHTCTRCVGPSEECIHCAKSFHQDWKVPACGEGFYDEMPGLPHKV 865

QY 159 CMKKGKTC-----GFKRG-----FETVRDILQHPSAKGNKGLCPPTSETRTIVORK 206

DB 866 CRRCDENCLSCGSSRNCRCAGTQLTGTCITNHTCSNADETFCEMVKSNRLC--BRK 923

RESULT 13

G02428
subtilisin-like proprotein convertase (EC 3.4.21.-) 5 precursor - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Mar-2000
C:Accession: G02428
R:Reudelhuber, T.L.
submitted to the EMBL Data Library, February 1996
A:Reference number: H01242
A:Accession: G02428
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-899 <PRU>
A:Cross-references: EMBL:U49114; NID:gl218057; PIDN:AAA91807.1; PID:gl218058
C:Genetics:
A:Gene: PCS
C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
C:Keywords: hydrolase; serine proteinase
P:148-386/Domain: subtilisin homology <SET>

Query Match 10.0%; Score 154; DB 2; Length 899;
Best Local Similarity 28.2%; Pred. No. 0.0089;
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;

QY 35 PNVISQ-GCOG-GCATCSQDNGCLSKPRLFFVLERIGMKQIGVCLSSCPGYYGTRYDPI 92

DB 624 PECSEVCGDGPDPHCND---CL-----HYTYKLR-----NNTRICVSSCPGPHY---HADK 669

QY 93 NKCTCKKVDCTCFNK--NFCTCKSGFYHL--LGKCLDSCPEGLANNHTMECVSIVHC 148

DB 670 KRCKKAPNCESCFGSHGDCMSCKGYFLNEETNSCVTHCPDGSYQDTKNLC----- 723

QY 149 EASESWSPWSPCMKKGKTC-GFKRGTYETVRDILQHPSAKGNKGLCPPTSE 197

DB 724 -----RKSENCKTCTEFHNTCEKRDGLSLQ-----GSRCSVSCE 758

RESULT 14

JC6148
subtilisin-like proprotein convertase (EC 3.4.21.-) homolog - human
N:Alternate names: PC6A protease
C:Species: Homo sapiens (man)
C>Date: 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 22-Jun-1999
C:Accession: JC6148
R:Miranda, L.; Wolf, J.; Pichuanes, S.; Duke, R.; Franzusoff, A.
Proc. Natl. Acad. Sci. U.S.A. 93, 7695-7700, 1996
A:Title: Isolation of the human PC6 gene encoding the putative host protease for HIV-1
A:Reference number: JC6148; MUID:96353880; PMID:8755538
A:Contents: CEM T-cell
A:Accession: JC6148
A:Molecule type: mRNA
A:Residues: 1-915 <MIR>
A:Cross-references: GB:U56387; NID:gl498312; PIDN:RAC50643.1; PID:gl498313
C:Comment: This protein functions as a soluble enzyme within the secretory pathway. It is
C:Genetics:
A:Gene: pc6A
C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
C:Keywords: glycoprotein; hydrolase; serine proteinase
P:164-402/Domain: subtilisin homology <SET>

Query Match 10.0%; Score 154; DB 2; Length 915;
Best Local Similarity 28.2%; Pred. No. 0.009;
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;

QY 35 PNVISQ-GCOG-GCATCSQDNGCLSKPRLFFVLERIGMKQIGVCLSSCPGYYGTRYDPI 92

DB 640 PECSEVCGDGPDPHCND---CL-----HYTYKLR-----NNTRICVSSCPGPHY---HADK 685

QY 93 NKCTCKKVDCTCFNK--NFCTCKSGFYHL--LGKCLDSCPEGLANNHTMECVSIVHC 148

DB 686 KRCKKAPNCESCFGSHGDCMSCKGYFLNEETNSCVTHCPDGSYQDTKNLC----- 739

QY 149 EASESWSPWSPCMKKGKTC-GFKRGTYETVRDILQHPSAKGNKGLCPPTSE 197

DB 740 -----RKSENCKTCTEFHNTCEKRDGLSLQ-----GSRCSVSCE 774

RESULT 15

A33837
insulin-like growth factor I receptor precursor - rat
N:Contains: insulin-like growth factor I alpha chain; insulin-like growth factor I beta
C:Species: Rattus norvegicus (Norway rat)
C>Date: 16-Mar-1990 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
C:Accession: JC2461; A33837; PC1131
R:Pedrini, M.T.; Giorgino, F.; Smith, R.J.
Biochem. Biophys. Res. Commun. 202, 1038-1046, 1994
A:Title: cDNA cloning of the rat IGF I receptor: Structural analysis of rat and human IGI
A:Reference number: JC2461; MUID:94324926; PMID:8048916
A:Accession: JC2461
A:Molecule type: mRNA
A:Residues: 1-1371 <PED>
R:Warner, H.; Woloschak, M.; Adamo, M.; Shen-Orr, Z.; Roberts Jr., C.T.; Leroith, D.
Proc. Natl. Acad. Sci. U.S.A. 86, 7451-7455, 1989
A:Title: Developmental regulation of the rat insulin-like growth factor I receptor gene.
A:Reference number: A33837; MUID:90017496; PMID:2477843
A:Accession: A33837
A:Molecule type: mRNA
A:Residues: 1-364 <WER>
A:Cross-references: GB:M27293
R:Kurachi, H.; Jobo, K.; Ohta, M.; Kawasaki, T.; Itoh, N.
Biochem. Biophys. Res. Commun. 187, 934-939, 1992
A:Title: A new member of the insulin receptor family, insulin receptor-related receptor,
A:Reference number: PC1130; MUID:92412145; PMID:1530648
A:Accession: PC1131

Query Match 10.0%; Score 154; DB 2; Length 899;
Best Local Similarity 28.2%; Pred. No. 0.0089;
Matches 48; Conservative 23; Mismatches 57; Indels 42; Gaps 11;

QY 35 PNVISQ-GCOG-GCATCSQDNGCLSKPRLFFVLERIGMKQIGVCLSSCPGYYGTRYDPI 92

DB 624 PECSEVCGDGPDPHCND---CL-----HYTYKLR-----NNTRICVSSCPGPHY---HADK 669

QY 93 NKCTCKKVDCTCFNK--NFCTCKSGFYHL--LGKCLDSCPEGLANNHTMECVSIVHC 148

DB 670 KRCKKAPNCESCFGSHGDCMSCKGYFLNEETNSCVTHCPDGSYQDTKNLC----- 723

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 16:54:24 ; Search time 8.46481 Seconds
(without alignments)
1716.232 Million cell updates/sec

Title: US-09-894-912A-32

Perfect score: 1535

Sequence: 1 MHLRLSCFFILNFMFYIG.....QKRRARDKQKSVSVSTVH 279

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	187	12.2	1877	1	PCK5 MOUSE
2	181.5	11.8	1696	1	PCK5 BRACL
3	179	11.7	969	1	PACA HUMAN
4	174	11.3	1679	1	FUR2 DROME
5	164	10.7	1877	1	PCK5 RAT
6	162	10.6	943	1	BL14 CAREL
7	161	10.5	937	1	PACA RAT
8	154	10.0	913	1	PCK5 HUMAN
9	153.5	10.0	374	1	WIFI_XENLA
10	149	9.7	1370	1	IGIR RAT
11	145.5	9.5	803	1	FSPO XENLA
12	144.5	9.4	2471	1	NTC2 RAT
13	142.5	9.3	807	1	FSPO RAT
14	142	9.3	1373	1	IGIR MOUSE
15	140.5	9.2	2470	1	NTC2 MOUSE
16	140	9.1	378	1	WIFI BRARE
17	139.5	9.1	1367	1	IGIR HUMAN
18	139	9.1	2524	1	NOTC XENLA
19	138.5	9.0	379	1	WIFI HUMAN
20	136	8.9	379	1	WIFI MOUSE
21	136	8.9	687	1	VS41_GIALA
22	133.5	8.7	2482	1	VWP PIG
23	131	8.5	1297	1	IRR HUMAN
24	131	8.5	2146	1	INSR DROME
25	131	8.5	2471	1	NTC2 HUMAN
26	130.5	8.5	1382	1	INSR HUMAN
27	130	8.5	2531	1	NTC1 MOUSE
28	130	8.5	2531	1	NTC1 RAT
29	129.5	8.4	1372	1	INSR MOUSE
30	129.5	8.4	1383	1	INSR RAT
31	129	8.4	2437	1	NTC1 BRARE
32	129	8.4	2556	1	NTC1 HUMAN
33	128	8.3	461	1	TR1A_PIG

ALIGNMENTS

RESULT 1

ID	PCK5_MOUSE	STANDARD;	PRT;	1877 AA.
AC	Q04592; Q62040;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Protein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)			
DE	(Protein convertase PCS) (Subtilisin/kexin-like protease PC5) (PC6)			
DE	(Subtilisin-like proprotein convertase 6) (SPC6).			
GN	PCS5.			
OS	Mus musculus (Mouse).			
OC	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE OF 330-1877 FROM N.A. (ISOFORM PCSB).			
RC	STRAIN=ICR; TISSUE=Intestine;			
RX	MEDLINE=93327934; PubMed=8335106;			
RA	Nakagawa T., Murakami K., Nakayama K.;			
RT	"Identification of an isoform with an extremely large Cys-rich region of PC6, a Kex2-like processing endoprotease.";			
RL	FEBS Lett. 327:165-171(1993).			
RN	[2]			
RP	SEQUENCE FROM N.A. (ISOFORM PCSA).			
RC	TISSUE=Brain, and Intestine;			
RX	MEDLINE=93224489; PubMed=8468318;			
RA	Nakagawa T., Hosaka M., Torii S., Watanabe T., Murakami K., Nakayama K.;			
RT	"Identification and functional expression of a new member of the mammalian Kex2-like processing endoprotease family: its striking structural similarity to PACE4.";			
RL	J. Biochem. 113:132-135(1993).			
RN	[3]			
RP	SEQUENCE FROM N.A. (ISOFORM PCSA).			
RC	TISSUE=Adrenal cortex;			
RX	MEDLINE=93342056; PubMed=8341687;			
RA	Lusson J., Vieau D., Hamelin J., Day R., Chretien M., Seidah N.G.;			
RT	"CDNA structure of the mouse and rat subtilisin/kexin-like PC5: a candidate proprotein convertase expressed in endocrine and nonendocrine cells.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).			
RN	[4]			
RP	PARTIAL SEQUENCE, AND SUBCELLULAR LOCATION.			
RC	MEDLINE=97103178; PubMed=8947550;			
RX	De Bie I., Marcinkiewicz M., Malide D., Lazure C., Nakayama K., Bendayan M., Seidah N.G.;			
RA	"The isoforms of proprotein convertase PC5 are sorted to different subcellular compartments.";			
RL	J. Cell Biol. 135:1261-1275(1996).			
RN	[5]			
RP	DEVELOPMENTAL EXPRESSION.			
RC	MEDLINE=96293359; PubMed=8698813;			
RX	Constam D.B., Calfon M., Robertson E.J.;			
RA	"SPC4, SPC6, and the novel protease SPC7 are coexpressed with bone morphogenetic proteins at distinct sites during embryogenesis.";			

34	127	8.3	1300	1	IRR_MOUSE	Q9wt14	mus musculus
35	124.5	8.1	667	1	TS11_GIALA	Q03185	giardia lam
36	124	8.1	581	1	IRR_RAT	Q64716	rattus norv
37	124	8.1	1300	1	IRR_CAVPO	P14617	cavia porce
38	123.5	8.0	375	1	DEK_HUMAN	P35659	homo sapien
39	123.5	8.0	1477	1	HTK7_HYDAT	Q25197	hydra atten
40	123	8.0	2871	1	FBN1_HUMAN	P35555	homo sapien
41	120.5	7.9	2871	1	FBN1_PIG	Q9tv36	sus scrofa
42	119	7.8	1713	1	LTBL_MOUSE	Q8c919	mus musculus
43	118.5	7.7	833	1	SRC2_MOUSE	P59222	mus musculus
44	118.5	7.7	1367	1	LT23_CAEEL	P24348	caenorhabdi
45	118	7.7	1170	1	TSP2_BOVIN	Q95116	bos taurus

RL J. Cell Biol. 134:181-191(1996).
 RN [6]
 RP DEVELOPMENTAL EXPRESSION.
 RX MEDLINE=97436919; PubMed=9291583;
 RA Rancourt S.L., Rancourt D.E.;
 RT "Murine subilisin-like proteinase SPC6 is expressed during embryonic
 RL implantation, somitogenesis, and skeletal formation";
 Dev. Genet. 21:75-81(1997).
 CC -!- FUNCTION: Likely to represent a widespread endoprotease activity
 CC within the constitutive and regulated secretory pathway. Capable
 CC of cleavage at the RX(K/R)R consensus motif. May be responsible
 CC for the maturation of gastrointestinal peptides. May be involved
 CC in the cellular proliferation of adrenal cortex via the activation
 CC of growth factors.
 CC -!- CATALYTIC ACTIVITY: Release of mature proteins from their
 CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa
 CC can be any amino acid and Yaa is Arg or Lys.
 CC -!- SUBCELLULAR LOCATION: PCSA IS SECRETED THROUGH THE REGULATED
 CC SECRETORY PATHWAY. PCSB IS A TYPE I MEMBRANE PROTEIN LOCALIZED TO
 CC A PARANUCLEAR POST-GOLGI NETWORK COMPARTMENT IN COMMUNICATION WITH
 CC EARLY ENDOSOMES.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=PCSB; Synonyms=Long;
 CC IsoId=Q04592-1; Sequence=Displayed;
 CC Name=PCSA; Synonyms=Short;
 CC IsoId=Q04592-2; Sequence=VSP_005438;
 CC -!- TISSUE SPECIFICITY: PCSA is expressed in most tissues but is most
 CC abundant in the intestine and adrenals. PCSB is expressed in the
 CC intestine, adrenals and lung but not in the brain.
 CC -!- DEVELOPMENTAL STAGE: Weakly expressed throughout the embryo,
 CC except in the developing nervous system, the ribs and the liver,
 CC but markedly upregulated at discrete sites during development. At
 CC E6.5, prominent expression observed in differentiated decidua. At
 CC E7.5, intense expression in extraembryonic endoderm, amnion and
 CC nascent mesoderm. At E8.5, abundant expression in somites and yolk
 CC sac followed by a confinement to dermamyotome compartment. Between
 CC E9.5 and E11.5, abundant expression in AER (thickened ectodermal
 CC cells of limb buds). At E12.5, expression in the limbs is confined
 CC to the condensing mesenchym surrounding the cartilage. At this
 CC stage, strong expression also detected in vertebral and facial
 CC cartilage primordia and in the muscle of the tongue. At E16.5
 CC abundant expression in epithelial cells of the intestinal villi.
 CC Isoform A is most abundant at all stages but significant levels of
 CC Isoform B occur at E12.5.
 CC -!- DOMAIN: The propeptide domain acts as an intramolecular chaperone
 CC assisting the folding of the zymogen within the endoplasmic
 CC reticulum.
 CC -!- DOMAIN: AC 1 and AC 2 (clusters of acidic amino acids) contain
 CC sorting information. AC 1 directs TGN localization and interacts
 CC with the TGN sorting protein PACS-1.
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC -!- SIMILARITY: Contains 1 homo B/P domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D17583; BAA04507.1; -;
 DR EMBL; D12619; BAA02143.1; -;
 DR EMBL; L14932; AAA74636.1; -;
 DR PIR; A48225; A48225.
 DR PIR; S34583; S34583.
 DR HSRP; Q99405; IMET.
 DR MEROPS; S08.076; -;
 DR MGD; MGI:97515; Pcsk5.
 DR InterPro; IPR006212; Furin repeat.
 DR InterPro; IPR009030; Grow_fac_recep.

DR InterPro; IPR000209; Peptidase_S8.
 DR InterPro; IPR002884; Peptidase_S8B.
 DR InterPro; IPR009020; Protease_inhib.
 DR Pfam; PF01483; P_protease_1.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR ProDom; PD000717; P_domain; 1.
 DR SMART; SM00261; FU; 22.
 DR PROSITE; PS00136; SUBTILASE ASP; 1.
 DR PROSITE; PS00137; SUBTILASE HIS; 1.
 DR PROSITE; PS00138; SUBTILASE SER; 1.
 DR Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
 KW Cleavage on pair of basic residues; Repeat; Alternative splicing;
 KW Transmembrane.
 FT SIGNAL 1 34
 FT PROPEP 35 116
 FT CHAIN 117 1877
 FT TYPE 5.
 FT PROTEIN CONVERTASE SUBTILISIN/KEXIN
 FT DOMAIN 117 1768 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1769 1789 POTENTIAL.
 FT DOMAIN 1790 1877 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 117 452 CATALYTIC.
 FT DOMAIN 464 602 HOMO B.
 FT DOMAIN 638 1753 CYS-RICH MOTIF (CRM) REGION.
 FT DOMAIN 1825 1844 AC 1.
 FT DOMAIN 1856 1877 AC 2.
 FT SITE 116 117 CLEAVAGE (AUTO-) (BY SIMILARITY).
 FT SITE 521 523 CELL ATTACHMENT SITE (POTENTIAL).
 FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 214 214 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 388 388 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 383 383 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 667 667 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 754 754 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 804 804 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 854 854 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 951 951 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1220 1220 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1317 1317 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1523 1523 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1711 1711 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1733 1733 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPPLIC 878 915
 FT GYDIDQGHGCTCEASCAKCGWGTQEDCISCPTVRVLD ->
 FT ATESWARGGFCMLVKNNLCKRKVLQQLCKCTCFQG
 FT (in isoform PCSA).
 FT /FTId=VSP_005438.
 FT Missing (in isoform PCSA).
 FT /FTId=VSP_005439.
 FT VARSPLIC 916 1877
 FT SEQUENCE 1877 AA; 209287 MW; EC850E2DF20BA1C3 CRC64;
 Query Match 12.2%; Score 187; DB 1; Length 1877;
 Best Local Similarity 27.8%; Pred. No. 1.2e-05;
 Matches 52; Conservative 21; Mismatches 72; Indels 42; Gaps 11;
 QY 38 SQCCQGGCATCSYNG--CLSKPRLFFVLERIGMKQIGVCLSSCSGGYGYTRPDI--N 93
 DB 1209 NQPCSSCKTC---NGSLCASCPTGMVLMQ-----ACVPCPQ3-----TPWSTSG 1253
 QY 94 KCTKCKVDCTCPKKNFCTCKKS---GFVLHGLKLDSCPEGLEANNHTMBCVSIHVCE 149
 DB 1254 SCEKSEDCVSCSGADLCQCCLSQPDNTLLHGRCTVHSCPEGYAKGVCE-----HCS 1308
 QY 150 ASEWSPWSPCKMKKGTGPKRGRTETVRDIL-----QHPSAKGMKMLCPPTSETRTIV 203
 DB 1309 ----SPCKTCEGNATSCNSCGDFVLDHGVCKWCTCEKHAIVEGVCKHCF--ERCQCIH 1362
 QY 204 QR--KKC 208
 DB 1363 EKTCKEC 1369

RESULT 2
ID PKCS_BRACL STANDARD; PRT; 1696 AA.
AC Q9NJ15; Q9NJ14; Q9NJ16;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE (Protein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)
DE (Protein convertase PCG-like) (aPC6).
PC6.
OS Branchiostoma californiensis (California lancelet) (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
NCBI_TaxID=7738;
[1]
RN
RP SEQUENCE FROM N.A. (ISOFORMS A; B AND C).
RA MEDLINE=20175281; PubMed=10708868;
RA Oliva A.A. Jr., Chan S.J., Steiner D.P.;
RT homologue of PC6 in the protochordate amphioxus.";
RL Biochim. Biophys. Acta 1477:338-348(2000).
CC -1- FUNCTION: likely to represent a widespread endoprotease activity
within the constitutive and regulated secretory pathway. Capable
of cleavage at the RX(K/R)R consensus motif (By similarity).
CC -1- CATALYTIC ACTIVITY: Release of mature proteins from their
proteins by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa
can be any amino acid and Yaa is Arg or Lys.
CC -1- SUBCELLULAR LOCATION: ISOFORM A AND ISOFORM C ARE SECRETED.
CC ISOFORM B IS A TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS;
CC Event=Alternative splicing; Named isoforms=3;
CC Name=B;
CC IsoId=Q9NJ15-1; Sequence=Displayed;
CC Name=A;
CC IsoId=Q9NJ15-2; Sequence=VSP_005444, VSP_005445;
CC Name=C;
CC IsoId=Q9NJ15-3; Sequence=VSP_005442, VSP_005443;
CC -1- DOMAIN: The propeptide domain acts as an intramolecular chaperone
assisting the folding of the zymogen within the endoplasmic
reticulum.
CC -1- SIMILARITY: Belongs to peptidase family S8.
CC -1- SIMILARITY: Contains 1 homo B/P domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>
or send an email to license@ebi.ac.uk).

DR EMBL; AF184615; AAF26300.1; -;
DR EMBL; AF184616; AAF26301.1; -;
DR EMBL; AF184617; AAF26302.1; -;
DR HSP; Q99405; IMPT.
DR MEROPS; S08.00B; -;
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow fac recep.
DR InterPro; IPR00209; Peptidase_S8.
DR InterPro; IPR002884; Peptidase_S8B.
DR InterPro; IPR009020; Protease_inhib.
DR Pfam; PF01483; P_protein; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR ProDom; PD000717; P_domain; 1.
DR SMART; SM00261; FU; 17.
DR PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
KW Cleavage on pair of basic residues; Repeat; Alternative splicing;
KW Transmembrane.
FT SIGNAL 1 25 POTENTIAL.

FT PROPEP 26 110
FT CHAIN 111 1696
FT
FT
FT DOMAIN 111 1618
FT TRANSMEM 1619 1639
FT DOMAIN 1640 1696
FT
FT DOMAIN 111 488
FT
FT DOMAIN 496 637
FT
FT DOMAIN 664 1649
FT SITE 110 111
FT ACT_SITE 192 192
FT ACT_SITE 233 233
FT ACT_SITE 407 407
FT CARBOHYD 246 246
FT CARBOHYD 529 529
FT CARBOHYD 885 885
FT VARSPLIC 1259 1323
FT
FT
FT
FT
FT VARSPLIC 1324 1696
FT
FT
FT VARSPLIC 1288 1343
FT
FT
FT
FT
FT VARSPLIC 1344 1696
FT
FT
FT
SQ SEQUENCE 1696 AA; 188410 MW; 281CB81784257CBD CRC64;

Query Match 11.8%; Score 181.5; DB 1; Length 1696;
Best Local Similarity 23.9%; Pred. No. 2.6e-05;
Matches 68; Conservative 23; Mismatches 72; Indels 121; Gaps 15;

Qy 33 MHPNVSQGGCGATCS--DYNGCLSCP----- 59
Db 1328 LHGDCDCSCHRECKTCDGPHDNCSCQPGSYLNDQCCSTHCPGTFEYDDSGEVL 1387
Qy 60 --RLFFV-----LERIGMKQIGVCLSSCPSPGYGTRYPDI-NKCTCKV 100
Db 1388 QCLCHVNCVTCGEGBEDCHCANDIKYQDGRVCVTECQEGH---YPLUTNECQCWS 1443
Qy 101 DCDTFC--NRNPTCKKSGFYHLGKCLDSCPEGL-----EANN 137
Db 1444 DCEFDGPRNDQCVTCFNYLYLVGKCLEDCPEGYDTMRQECGCHPSCATCNCGN 1503
Qy 138 HTWECVSIHV-----CRASEW-----SPWSPCKKGTGCFKGTETVRDIL 180
Db 1504 Y--NCLSCPYSGLGBCGVCYPMCEHEHYVETKQICECDNSCKTC---RG--STAHDCL 1556
Qy 181 QHPSAKG---KGMLC-----PTSE-----TRTCIVQRK 206
Db 1557 SCAPYGYHAMKHLCTACCEGSPENYCCICHESTRLCTIDRE 1600

RESULT 3
ID PAC4 HUMAN STANDARD; PRT; 969 AA.
AC P29122; Q15099; Q15100; Q9UEG7; Q9UEJ1; Q9UEJ2; Q9UEJ7; Q9UEJ8;
AC Q9UEJ9; Q9Y4G9; Q9Y4H0; Q9Y4H1;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Paired basic amino acid cleaving enzyme 4 precursor (EC 3.4.21.-)
DE (Subtilisin/kexin-like protease FACE4) (Subtilisin-like proprotein
convertase 4) (SPC4).
GN FACE4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;

[1] SEQUENCE FROM N.A. (ISOFORMS PACE4A-I AND PACE4B).

RP TISSUE=Hepatoma, and kidney;

RC MEDLINE=92075167; PubMed=1741956;

RA Klefer M.C., Tucker J.E., Joh R., Landsberg K.E., Saltman D.,

RA Barr P.J.;

RT "Identification of a second human subtilisin-like protease gene in

RT the fes/fps region of chromosome 15.";

RL DNA Cell Biol. 10:757-769(1991).

[2] SEQUENCE FROM N.A. (ISOFORMS PACE4C AND PACE4D).

RP TISSUE=Placenta;

RC MEDLINE=94235049; PubMed=8179631;

RA Tsuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,

RA Matsuda Y.;

RT "Identification of novel cDNAs encoding human kexin-like protease,

RT PACE4 isoforms.";

RL Biochem. Biophys. Res. Commun. 200:943-950(1994).

[3] HRRATUM.

RP MEDLINE=95071480; PubMed=7980617;

RA Tsuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,

RA Matsuda Y.;

RT "Identification of novel cDNAs encoding human kexin-like protease,

RT PACE4 isoforms.";

RL Biochem. Biophys. Res. Commun. 204:1381-1382(1994).

[4] SEQUENCE FROM N.A. (ISOFORM PACE4A-II).

RP TISSUE=Placenta;

RC MEDLINE=98021085; PubMed=9378725;

RA Mori K., Imamaki A., Kii S., Nagamune H., Nagahama M., Tsuji A.,

RA Matsuda Y.;

RT "Identification of a novel PACE4 isoform, PACE4B.";

RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.

[5] SEQUENCE FROM N.A. (ISOFORMS PACE4E-I AND PACE4E-II).

RP TISSUE=Cerebellum;

RC MEDLINE=97335942; PubMed=9192737;

RA Mori K., Kii S., Tsuji A., Nagahama M., Imamaki A., Hayashi K.,

RA Akamatsu T., Nagamune H., Matsuda Y.;

RT "A novel human PACE4 isoform, PACE4E is an active processing protease

RT containing a hydrophobic cluster at the carboxy terminus.";

RL J. Biochem. 121:941-948(1997).

[6] SEQUENCE FROM N.A. (ISOFORMS PACE4A-I; PACE4A-II; PACE4C; PACE4D;

RP PACE4E-I AND PACE4E-II).

RX MEDLINE=98021085; PubMed=9378725;

RA Tsuji A., Hine C., Tamai Y., Yonemoto K., Mori K., Yoshida S.,

RA Bando M., Sakai E., Mori K., Akamatsu T., Matsuda Y.;

RT "Genomic organization and alternative splicing of human PACE4 (SPC4),

RT kexin-like processing endoprotease.";

RL J. Biochem. 122:438-452(1997).

[7] ALTERNATIVE SPLICING (ISOFORM PACE4CS).

RP MEDLINE=97064242; PubMed=8906861;

RX Zhong M., Benjannet S., Lazure C., Munzer S., Seidah N.G.;

RT "Functional analysis of human PACE4-A and PACE4-C isoforms:

RT Identification of a new PACE4-CS isoform.";

RL FEBS Lett. 396:31-36(1996).

[8] CHARACTERIZATION.

RP MEDLINE=99233559; PubMed=10215603;

RA Sucic J.F., Moehring J.M., Innocencio N.M., Luchini J.W.,

RA Moehring T.J.;

RT "Endoprotease PACE4 is Ca²⁺-dependent and temperature-sensitive and

RT can partly rescue the phenotype of a furin-deficient cell strain.";

RL Biochem. J. 339:639-647(1999).

[9] PROCESSING.

RP MEDLINE=98408849; PubMed=9738469;

RX Nagahama M., Taniguchi T., Hashimoto E., Imamaki A., Mori K.,

RA Tsuji A., Matsuda Y.;

RT "Biothetic processing and quaternary interactions of proprotein

RT convertase SPC4 (PACE4).";

FEBS Lett. 434:155-159(1998).

-1- FUNCTION: Likely to represent an endoprotease activity within the

CC constitutive secretory pathway, with unique restricted

CC distribution in both neuroendocrine and non-neuroendocrine tissues

CC and capable of cleavage at the RX(K/R)R consensus motif.

CC -1- CATALYTIC ACTIVITY: Release of mature proteins from their

CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa

CC can be any amino acid and Yaa is Arg or Lys.

CC -1- COFACTOR: Calcium (Potential).

CC -1- SUBUNIT: The PACE4A-I precursor protein seems to exist in the

CC reticulum endoplasmic as both a monomer and a dimer-sized complex

CC whereas mature PACE4A-I exists only as a monomer, suggesting that

CC propeptide cleavage affects its tertiary or quaternary structure.

CC -1- SUBCELLULAR LOCATION: PACE4A-I and PACE4A-II are secreted. PACE4C

CC and PACE4CS are not secreted and remain probably in zymogen form

CC in endoplasmic reticulum. PACE4E-I and PACE4E-II are retained

CC intracellularly probably through a hydrophobic cluster in their C-

CC terminus. PACE4B might be secreted.

CC -1- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=8;

CC Name=PACE4A-I; Synonyms=PACE4;

CC IsoId=P29122-1; Sequence=Displayed;

CC Name=PACE4A-II;

CC IsoId=P29122-2; Sequence=VSP_005436;

CC Name=PACE4B; Synonyms=PACE4.1;

CC IsoId=P29122-3; Sequence=VSP_005428, VSP_005429;

CC Note=Probably enzymatically inactive;

CC Name=PACE4C;

CC IsoId=P29122-4; Sequence=VSP_005432, VSP_005433;

CC Note=Probably enzymatically inactive;

CC Name=PACE4CS;

CC IsoId=P29122-5; Sequence=VSP_005430, VSP_005431;

CC Note=Probably enzymatically inactive;

CC Name=PACE4D;

CC IsoId=P29122-6; Sequence=VSP_005427, VSP_005434, VSP_005435;

CC Note=Probably enzymatically inactive;

CC Name=PACE4E-I;

CC IsoId=P29122-7; Sequence=VSP_005437;

CC Name=PACE4E-II;

CC IsoId=P29122-8; Sequence=VSP_005436, VSP_005437;

CC TISSUE SPECIFICITY: Each PACE4 isoform exhibits a unique

CC restricted distribution. PACE4A-I is expressed in heart, brain,

CC placenta, lung, skeletal muscle, kidney, pancreas, but at

CC comparatively higher levels in the liver. PACE4A-II is at least

CC expressed in placenta. PACE4B was only found in the embryonic

CC kidney cell line from which it was isolated. PACE4C and PACE4D are

CC expressed in placenta. PACE4E-I is expressed in cerebellum,

CC placenta and pituitary. PACE4E-II is at least present in

CC cerebellum.

CC -1- DOMAIN: The propeptide domain acts as an intramolecular chaperone

CC assisting the folding of the zymogen within the endoplasmic

CC reticulum. Isoform PACE4D lacks the propeptide domain.

CC -1- SIMILARITY: Belongs to peptidase family S8.

CC -1- SIMILARITY: Contains 1 homo B/P domain.

CC -1- SIMILARITY: Contains 1 PLAC domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.iesb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC -----

CC EMBL; M80482; AAA5998.1; --

CC EMBL; AB001914; BAA21620.1; --

CC EMBL; AB001898; BAA21620.1; JOINED.

CC EMBL; AB001900; BAA21620.1; JOINED.

CC EMBL; AB001901; BAA21620.1; JOINED.

CC EMBL; AB001902; BAA21620.1; JOINED.

CC EMBL; AB001903; BAA21620.1; JOINED.

CC EMBL; AB001904; BAA21620.1; JOINED.

CC EMBL; AB001905; BAA21620.1; JOINED.

```

DR EMBL; AB001914; BAA21621.1; -.
DR EMBL; AB001898; BAA21621.1; JOINED.
DR EMBL; AB001900; BAA21621.1; JOINED.
DR EMBL; AB001901; BAA21621.1; JOINED.
DR EMBL; AB001902; BAA21621.1; JOINED.
DR EMBL; AB001903; BAA21621.1; JOINED.
DR EMBL; AB001904; BAA21621.1; JOINED.
DR EMBL; AB001905; BAA21621.1; JOINED.
DR EMBL; AB001906; BAA21621.1; JOINED.
DR EMBL; AB001907; BAA21621.1; JOINED.
DR EMBL; AB001908; BAA21621.1; JOINED.
DR EMBL; AB001909; BAA21621.1; JOINED.
DR EMBL; AB001914; BAA21622.1; -.
DR EMBL; AB001901; BAA21622.1; JOINED.
DR EMBL; AB001902; BAA21622.1; JOINED.
DR EMBL; AB001903; BAA21622.1; JOINED.
DR EMBL; AB001904; BAA21622.1; JOINED.
DR EMBL; AB001905; BAA21622.1; JOINED.
DR EMBL; AB001906; BAA21622.1; JOINED.
DR EMBL; AB001907; BAA21622.1; JOINED.
DR EMBL; AB001908; BAA21622.1; JOINED.
DR EMBL; AB001914; BAA21623.1; -.
DR EMBL; AB001898; BAA21623.1; JOINED.
DR EMBL; AB001900; BAA21623.1; JOINED.
DR EMBL; AB001901; BAA21623.1; JOINED.
DR EMBL; AB001902; BAA21623.1; JOINED.
DR EMBL; AB001903; BAA21623.1; JOINED.
DR EMBL; AB001904; BAA21623.1; JOINED.
DR EMBL; AB001905; BAA21623.1; JOINED.
DR EMBL; AB001906; BAA21623.1; JOINED.
DR EMBL; AB001907; BAA21623.1; JOINED.
DR EMBL; AB001908; BAA21623.1; JOINED.
DR EMBL; AB001914; BAA21624.1; -.
DR EMBL; AB001898; BAA21624.1; JOINED.
DR EMBL; AB001900; BAA21624.1; JOINED.
DR EMBL; AB001901; BAA21624.1; JOINED.
DR EMBL; AB001902; BAA21624.1; JOINED.
DR EMBL; AB001903; BAA21624.1; JOINED.
DR EMBL; AB001904; BAA21624.1; JOINED.

Query Match      11.7%; Score 179; DB 1; Length 969;
Best Local Similarity 23.3%; Pred. No. 2.3e-05;
Matches 56; Conservative 23; Mismatches 89; Indels 72; Gaps 12;

Oy 18 YIGSONASRGRRQRMRHNVSGGGGCGATCSD--YNGCLSKCKRLPFLVLRIGMKQIGV 75
Db 737 YFGDTRARCR-----CHKGCETSSRAATQCLSCR-RCFY-----HHQEWNT 779
Oy 76 CLSSCPGYYGTRYPDINKTKCKVDCTCFNK-NFTCKCKSGFYHLHGKCLDSCEGLE 134
Db 780 CVTLCPAGFYADE--SQKNCLKCHPSCKKCVDEPEKCTVCKGFSLARGSCITPCDCEPTY 837
Oy 135 ANNHTEC-----VSIVCEAS-----EWS-----PWSF 158
Db 838 FDSLLRGCECHTGTGTCVGPREECHCAKNFHFHDKWCPACGEGYPPEMGLPHKV 897
Oy 159 CMKKGKTC---GPKRG-----TETVRDILQHPSAKGKGNLCPTSETRTCTIVQRK 206
Db 398 CRRDCNCLSCAGSRNCRCKTGTGTQLTGTCITNHTCSNADETCFEMVKNRLC--ERK 955

RESULT 4
FUR2 DROME
AC P30432; Q24301; PRT; 1679 AA.
DT 01-APR-1993 (Rel. 25, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
Db Furin-like protease 2 precursor (BC 3.4.21.75) (Furin 2).
GN FUR2 OR CQ18734/CG4235.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

```

```

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R, Tuebingen, and Iso-1;
RX MEDLINE=92381036; PubMed=1512259;
RA Roebroek A.J.M., Creemers J.W.M., Pauli I.G.L., Kurzik-Dumke U.,
RA Rentsop M., Gatteff E.A.F., Leunissen J.A.M., van de Ven W.J.M.;
RA "Cloning and functional expression of Dfurin2, a subtilisin-like
RT proprotein processing enzyme of Drosophila melanogaster with multiple
RT repeats of a cysteine motif.";
RL J. Biol. Chem. 267:17208-17215(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ISO-1;
RX MEDLINE=95186060; PubMed=7880443;
RA Roebroek A.J.M., Ayoubi T.A.Y., Creemers J.W.M., Pauli I.G.L.,
RA van de Ven W.J.M.;
RA "The Dfur2 gene of Drosophila melanogaster: genetic organization,
RT expression during embryogenesis, and pro-protein processing activity
RT of its translational product Dfurin2.";
RL DNA Cell Biol. 14:233-234(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.B.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Beremand B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fodor C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.B., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasbo P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier B., Spradling A.C., Turner R., Venter E., Wang A.H., Wang X.,
RA Swirskas R., Tector C.C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -1- FUNCTION: Furin is likely to represent the ubiquitous endoprotease
CC activity within constitutive secretory pathways and capable of
CC cleavage at the RX(R/R)R consensus motif (By similarity).
CC -1- CATALYTIC ACTIVITY: Release of mature proteins from their
CC proproteins by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa
CC can be any amino acid and Yaa is Arg or Lys. Releases albumin,
CC complement component C3 and von Willebrand factor from their
CC respective precursors.

```

CC -1- TISSUE SPECIFICITY: Transient expression in a subset of central
 CC nervous system neurons during embryonic stages 12-13. Expression
 CC in developing tracheal tree from stage 13 to end of embryonic
 CC development.
 CC -1- SIMILARITY: Belongs to peptidase family S8. Furin subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

DR EMBL; M94375; AAA28551.1; -;
 DR EMBL; L33831; AAA69860.1; -;
 DR EMBL; AE003502; AAF48598.1; -;
 DR PIR; A43434; A43434.
 DR HSSP; Q99405; IMPT.
 DR MEROPS; S08.049; -;
 DR FlyBase; FBgn0004598; Fur2.
 DR GO; GO:0004276; F: furin activity; IDA.
 DR InterPro; IPR006212; Furin repeat.
 DR InterPro; IPR009030; Grow fac recep.
 DR InterPro; IPR002029; Peptidase_S8.
 DR InterPro; IPR002884; Peptidase_S8B.
 DR InterPro; IPR009020; Protease_inhib.
 DR Pfam; PF01483; P_protease; 2.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR ProDom; PD000717; P_domain; 1.
 DR SMART; SM00261; FU_10.
 DR PROSITE; PS00136; SUBTILASE ASP; 1.
 DR PROSITE; PS00137; SUBTILASE HIS; 1.
 DR PROSITE; PS00138; SUBTILASE SER; 1.
 DR Hydrolase; Serine protease; Glycoprotein; Signal; Transmembrane;
 KW Multigene family; Zymogen; Repeat.
 FT SIGNAL; 1 2
 FT PROPEP 1 318 POTENTIAL.
 FT CHAIN 319 1679 FURIN-LIKE PROTEASE 2.
 FT ACT_SITE 417 417 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 456 456 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 637 637 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DOMAIN 961 1443 10 X TANDEN REPEATS, CYS-RICH.
 FT REPEAT 961 1006 1.
 FT REPEAT 1007 1056 2.
 FT REPEAT 1057 1103 3.
 FT REPEAT 1104 1152 4.
 FT REPEAT 1153 1204 5.
 FT REPEAT 1205 1253 6.
 FT REPEAT 1254 1298 7.
 FT REPEAT 1299 1345 8.
 FT REPEAT 1346 1392 9.
 FT REPEAT 1393 1443 10.
 FT TRANSMEM 1512 1532 POTENTIAL.
 FT DOMAIN 1533 1679 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 3 3 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 109 109 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 130 130 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 205 205 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 442 442 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 480 480 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 927 927 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1060 1060 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1181 1181 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1274 1274 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1277 1277 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1439 1439 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CONFLICT 152 153 MISSING (IN REF. 1).
 FT CONFLICT 177 177 V -> F (IN REF. 1).
 FT CONFLICT 213 213 V -> VDQL (IN REF. 1).
 SQ SEQUENCE 1679 AA; 183369 MW; 3F9E749F0B021CP6 CRC64;

Query Match 11.3%; Score 174; DB 1; Length 1679;
 Best Local Similarity 24.5%; Pred. No. 8.6e-05;
 Patches 48; Conservative 15; Mismatches 61; Indels 72; Gaps 6;
 QY 23 NASRRERQRMHFN---VSQGGCGGCATCSDY--NGCLSKCKPRLPFVLERIGMKQIGVCL 77
 DB 994 NTCVSRCPSPFPNQVGICWPCDHTCETCAGAGPDSCLTCAPAHLRVID-----LAVCL 1047
 QY 78 SSCPSGY-----GTRYPDI 92
 DB 1048 QFCPDGYFENSRNRTVCPCEPNCASQDHPETCTCDHLLVMHKKYCSACPLDTYETSD 1107
 QY 93 NKCTKCKVDCDTCF--NKNFCTCKSGFYHLGKCLDSCPEGLNANNHTWECVSIHVCEA 150
 DB 1108 NKCAFCHSTCATNGPFDQDICTRSRYAWQNKCLISCPDGFYADKKRLSCM----- 1160
 QY 151 SEWSPWSPCMKGGKTC 166
 DB 1161 -----PCQEGCKTC 1169
 RESULT 5
 PCS5 RAT
 ID PCS5 RAT STANDARD; PRT; 1877 AA.
 AC P41I3; Q62914;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)
 DE (Proprotein convertase PCS5) (Subtilisin/kexin-like protease PCS)
 DE (rPCS5) (PC6) (Fragment).
 GN PCS5.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM SHORT), AND TISSUE SPECIFICITY.
 RC TISSUE=Adrenal gland;
 RX MEDLINE=93342056; PubMed=8341687;
 RA Lussion J., Vieau D., Hamelin J., Day R., Chretien M., Seidah N.G.;
 RT "cDNA structure of the mouse and rat subtilisin/kexin-like PCS5; a
 RT candidate proprotein convertase expressed in endocrine and
 RT nonendocrine cells.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).
 RN [2]
 RP SEQUENCE OF 1676-1877 FROM N.A. (ISOFORM LONG).
 RC TISSUE=Adrenal gland;
 RA De Bie I., Marcinkiewicz M., Nakayama K., Lazure C., Seidah N.G.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SUBCELLULAR LOCATION.
 RX Xiang Y., Molloy S.S., Thomas L., Thomas G.;
 RT "The PC6B cytoplasmic domain contains two acidic clusters that direct
 RT sorting to distinct trans-Golgi network/endosomal compartments.";
 RL Mol. Biol. Cell 11:1257-1273(2000).
 RN [4]
 RP DEVELOPMENTAL EXPRESSION.
 RX MEDLINE=97166043; PubMed=9013936;
 RA Zheng M., Seidah N.G., Pintar J.E.;
 RT "The developmental expression in the rat CNS and peripheral tissues of
 RT proteases PC5 and PACE4 mRNAs: comparison with other protein
 RT processing enzymes.";
 RL Dev. Biol. 181:268-283(1997).
 CC -1- FUNCTION: Likely to represent a widespread endoprotease activity
 CC within the constitutive and regulated secretory pathway. Capable
 CC of cleavage at the RX(K/R)R consensus motif. May be responsible
 CC for the maturation of gastrointestinal peptides. May be involved
 CC in the cellular proliferation of adrenal cortex via the activation
 CC of growth factors.
 CC -1- CATALYTIC ACTIVITY: Release of mature proteins from their

proteins by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa can be any amino acid and Yaa is Arg or Lys.

-1- SUBCELLULAR LOCATION: PCSA IS SECRETED THROUGH THE REGULATED SECRETORY PATHWAY. PCSB IS A TYPE I MEMBRANE PROTEIN LOCALIZED TO A PARANUCLEAR POST-GOLGI NETWORK COMPARTMENT IN COMMUNICATION WITH EARLY ENDOSOMES (BY SIMILARITY).

-1- ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=2;
Comment=Additional isoforms seem to exist;
Name=PCSB; Synonyms=Long;
IsoId=P41413-1; Sequence=Displayed;
Name=PCSA; Synonyms=Short;
IsoId=P41413-2; Sequence=VSP_005440, VSP_005441;

-1- TISSUE SPECIFICITY: Expressed in the intestine, brain, adrenal gland, anterior pituitary, thyroid, ovaries, testis and lung. Highest levels are found in the gut, duodenum, jejunum and ileum. Expression is higher in female than in male reproductive organs.

-1- DEVELOPMENTAL STAGE: First detected at E9 in highly restricted regions of the neural tube, in caudal myotomes, and at the materno-embryonic junction of the uterus. At E10, restricted expression is detected in the optic and otic vesicles, the roof of midbrain, and trunk myotomes. By midgestation (E13-E16), expression in the developing nervous system has expanded to multiple regions including hippocampus, thalamus, hypothalamus, brain stem, and spinal cord. Expression is also detected in several peripheral organ systems, including gut, lung, adrenal and kidney primordia.

-1- DOMAIN: The propeptide domain acts as an intramolecular chaperone assisting the folding of the zymogen within the endoplasmic reticulum.

-1- DOMAIN: AC 1 and AC 2 (clusters of acidic amino acids) contain sorting information. AC 1 directs TGN localization and interacts with the TGN sorting protein PACS-1.

-1- SIMILARITY: Belongs to peptidase family S8.

-1- SIMILARITY: Contains 1 homo B/P domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

EMBL; L14933; AAA99906.1; --
EMBL; U47014; AAA87888.1; --
PIR; B48225; B48225.
HSP; Q99405; IMPT.
MEROPS; S08.076; --
InterPro; IPR006212; Purin repeat.
InterPro; IPR000209; Peptidase S8.
InterPro; IPR002884; Peptidase S8B.
InterPro; IPR009020; Protease Inhib.
Pfam; PF01483; P_protease; 1.
PRINTS; PR00723; SUBTILISIN.
ProDom; PD000717; P_domain; 1.
SMART; SM0261; FU; 6.
PROSITE; PS00136; SUBTILASE ASP; 1.
PROSITE; PS00137; SUBTILASE HIS; 1.
PROSITE; PS00138; SUBTILASE SER; 1.
Hydrolase; Serine protease; Zymogen; Signal;
Cleavage on pair of basic residues; Alternative splicing; Repeat;
Transmembrane.

SIGNAL 1 34 BY SIMILARITY.
FT PROPEP 35 116 BY SIMILARITY.
FT CHAIN 117 1877 PROPEPTIN CONVERTASE SUBTILISIN/KEXIN TYPE 5.
FT DOMAIN 117 1768 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1769 1789 POTENTIAL.
FT DOMAIN 1790 1877 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 117 452 CATALYTIC.
FT DOMAIN 464 602 HOMO B.
FT DOMAIN 638 1753 CYS-RICH MOTIF (CRM) REGION.

FT DOMAIN 1825 1844 AC 1.
FT DOMAIN 1856 1877 CLEAVAGE (AUTO-) (BY SIMILARITY).
FT SITE 116 117 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 521 523 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 214 214 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 383 383 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 667 667 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 754 754 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 804 804 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 854 854 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1710 1710 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1732 1732 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 878 915 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX -->
ATBESWAGGFCMLVKNNLQCRKVLQALCLCKCTCFQG
(in isoform PCSA).
FT FTid=VSP_005440.
FT VARSPLIC 916 1877 Missing (in isoform PCSA).
FT FTid=VSP_005441.
SQ SEQUENCE 1877 AA; 207888 MW; 890955DC6053444 CRC64;
Query Match 10.7%; Score 164; DB 1; Length 1877;
Best Local Similarity 27.7%; Pred. No. 0.00047;
Matches 48; Conservative 20; Mismatches 57; Indels 48; Gaps 9;
QY 35 PNVSG-GCOG---GCATCSYNGCLSCRPFLERIGMKIGVCLSCSPSGYGYTRY 89
DB 640 PECSEVGCDGPGDHCTDCLHYHVKLNTR-----ICVSSCPGHP---H 682
QY 90 PDINKTKKVCDCVCPKNP--CTKCKSGFYH--LGKLDSCPEGLNHNHTWCVSI 145
DB 683 ADKRCRCAPNCESCFGSHADCLSKYGYFLNEETSSCVAQPGSGYQDKKNIC--- 739
QY 146 VHCEASEWSPSCMKKGKTC-GPKRGTEVRDILQHPSAKGKGLCPPTSE 197
DB 740 -----GKSENCKTCTGFHNCTCKGGLSLQ-----GSRCSVTCE 774
RESULT 6
BL14 CAREL STANDARD; PRT; 943 AA.
AC P51559; O44762; O44763; O44764; O44765; O44766;
DT 01-OCT-1996 (Rel. 34, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endoprotease bli-4 precursor (EC 3.4.21.-) (Blisterase) (Blistered cuticle protein 4)
DE Bui-4 OR KPC-4 OR K04F10.4.
OS Caenorhabditis elegans.
OC Rukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A; B; C AND D), AND FUNCTION.
RC STRAIN=Br18tol N2;
RX MEDLINE=95293228; PubMed=7774813;
RT "The bli-4 locus of Caenorhabditis elegans encodes structurally distinct kex2/subtilisin-like endoproteases essential for early development and adult morphology.";
RL Genes Dev. 9:956-971(1995).
RN [2]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=Br18tol N2;
RA Latraille P., Wameley P.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN MUTAGENESIS.
RX MEDLINE=20363723; PubMed=10903434;
RA Thacker C., Srayko M., Rose A.M.;
RT "Mutational analysis of bli-4/kpc-4 reveals critical residues required

FT DISULFID 289 291 POTENTIAL.
 FT DISULFID 300 309 POTENTIAL.
 FT DISULFID 305 315 POTENTIAL.
 FT DISULFID 321 323 POTENTIAL.
 FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 240 240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 374 AA; 41071 MW; E26P973B0F00ACP8 CRC64;

Query Match 10.0%; Score 153.5; DB 1; Length 374;
 Best Local Similarity 24.4%; Pred. No. 0.00056;
 Matches 60; Conservative 23; Mismatches 78; Indels 85; Gaps 14;

QY 41 COGCGAT---CSYNGC-----LSCPELFFVLIRGMQIGVC-----LSSC 80
 DB 177 CTGCGRNGFGNDKHVCECPDGFYGPCEKALCMPRCM-----NGGLCVTPGLCIC 227
 QY 81 PSYGYGTRYPDINKTKCKVDCTCFNKNFCTCKSGFYHLGLKCLDSCPEGLNANNHTM 140
 DB 228 PPGYGINCDKVNCTHIC-LNGGTCTP-----YPGKCI--CPSGYEGE----- 266
 QY 141 ECYSIVHCRASEWSPWSPCKMKKGTCTGFKGTETVRDILQHPSAKG-KGNLC-----P 193
 DB 267 -----QCETSKQ--QPCRNKGKCSGKNK-----CKSKGYQDGLCSKPVCBP 307
 QY 194 PTSTFTCIWQRK-KCSGGERKKGRERKKLKKERKETSSTSSDSKGLSSTETPDQQ 252
 DB 308 SCGAHGTCIFBNKQCKQCKEGWNG-----RYCNKKYGSMLMALRPTGSRNRQHTSPK 359
 QY 253 ENKERQ 258
 DB 360 RTEDRQ 365

RESULT 10

IGIR RAT STANDARD; PRT; 1370 AA.
 AC P24052;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Insulin-like growth factor I receptor precursor (EC 2.7.1.112).
 GN IGF1R.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=95277910; PubMed=7758167;
 RA Du J., Delafontaine P.;
 RT "Inhibition of vascular smooth muscle cell growth through antisense
 transcription of a rat insulin-like growth factor I receptor cDNA.";
 RL Circ. Res. 76:963-972(1995).
 RN [2]
 RP SEQUENCE OF 1-364 FROM N.A.
 RC STRAIN=Sprague-Dawley;
 RX MEDLINE=90017496; PubMed=2477843;
 RA Werner H., Woloschak M., Adamo M., Shen-Orr Z., Roberts C.T. Jr.,
 RA Leroith D.;
 RT "Developmental regulation of the rat insulin-like growth factor I
 receptor gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:7451-7455(1989).
 RN [3]
 RP SEQUENCE OF 913-1017 FROM N.A.
 RX MEDLINE=92412145; PubMed=1530648;
 RA Kurachi H., Jobo K., Ohta M., Kawasaki T., Itoh N.;
 RT "A new member of the insulin receptor family, insulin
 receptor-related receptor, is expressed preferentially in the
 kidney.";
 RT Biochem. Biophys. Res. Commun. 187:934-939(1992).
 CC -!- FUNCTION: THIS RECEPTOR BINDS INSULIN-LIKE GROWTH FACTOR I (IGF I)
 WITH A HIGH AFFINITY AND IGF II WITH A LOWER AFFINITY. IT HAS A

CC CC TYROSINE-PROTEIN KINASE ACTIVITY.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -!- SUBUNIT: Tetramer of 2 alpha and 2 beta chains linked by disulfide
 CC bonds. The alpha chains contribute to the formation of the ligand-
 CC binding domain, while the beta chain carries the kinase domain.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- SIMILARITY: Belongs to the tyr family of protein kinases. Insulin
 CC receptor subfamily.
 CC -!- SIMILARITY: Contains 2 fibronectin type III domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L29232; AAA41392.1; -;
 CC EMBL; M27293; AAA41384.1; -;
 CC HSP; P06213; IIRK.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR008957; FN_III-like.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR006211; Furin-like.
 DR InterPro; IPR006212; Furin repeat.
 DR InterPro; IPR009030; Growth factor receptor.
 DR InterPro; IPR000719; Prot_kinase.
 DR InterPro; IPR002011; RecepttyrkinasII.
 DR InterPro; IPR001245; Tyr_kinase.
 DR InterPro; IPR008266; Tyr_kinase_AS.
 DR Pfam; PF00041; fn3; 2.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; kinase; 1.
 DR Pfam; PF01030; Recept_L domain; 2.
 DR PRINTS; PR0109; TYRKINASE.
 DR PRODOM; PD000001; Prot_kinase; 1.
 DR SMART; SM00060; FN3; 3.
 DR SMART; SM00261; FU; 1.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
 DR PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
 DR PROSITE; PS50011; PROTEIN KINASE_DOM; 1.
 KW Transferase; Tyrosine-protein kinase; Receptor; Transmembrane;
 KW Glycoprotein; ATP-binding; Phosphorylation; Repeat; Signal.
 FT SIGNAL 1 30
 FT CHAIN 31 741 INSULIN-LIKE GROWTH FACTOR I RECEPTOR,
 FT ALPHA-CHAIN.
 FT CHAIN 742 1370 INSULIN-LIKE GROWTH FACTOR I RECEPTOR,
 FT BETA-CHAIN.
 FT DOMAIN 742 936 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 937 960 POTENTIAL.
 FT DOMAIN 961 1370 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 608 829 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 830 929 FIBRONECTIN TYPE-III 2.
 FT NP_BIND 1006 1014 PROTEIN KINASE.
 FT BINDING 1034 1034 ATP (BY SIMILARITY).
 FT ACT_SITE 1136 1136 ATP (BY SIMILARITY).
 FT DISULFID 215 224 BY SIMILARITY.
 FT DISULFID 219 230 BY SIMILARITY.
 FT DISULFID 231 239 BY SIMILARITY.
 FT DISULFID 235 248 BY SIMILARITY.
 FT DISULFID 251 260 BY SIMILARITY.
 FT DISULFID 264 276 BY SIMILARITY.
 FT DISULFID 282 303 BY SIMILARITY.
 FT DISULFID 307 321 BY SIMILARITY.
 FT DISULFID 324 328 BY SIMILARITY.
 FT CARBOHYD 51 51 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 439 439 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 535 535 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 608 608 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 641 641 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 748 748 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 757 757 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 765 765 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 901 901 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 914 914 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT MOD RES 1166 1166 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CONFLICT 985 986 AD -> PY (IN REP. 3). CRC64;
 SQ SEQUENCE 1370 AA; 155395 MM; A5946897A41CB145 CRC64;
 Query Match 9.7%; Score 149; DB 1; Length 1370;
 Best Local Similarity 28.7%; Pred. No. 0.0039;
 Matches 47; Conservative 20; Mismatches 57; Indels 40; Gaps 11;
 QY 34 HPWVGCGCGGATCSDYNGCLCKPRLFFVLRIGMKQIGVCLSSCPGYYGTRYPDIN 93
 DB 232 HPE---CLGCHTPDINTTCVACHYY-----KGVCVPACPGTY--RP--- 271
 QY 94 KCTKVKDCTCFN-KNFCCKSGFYHLGKCLDSCEGLEANN-HTMECVSVHCEAS 151
 DB 272 EGWRC-VDRDFCANIPNABSSDSGFGVHDGECQBCPSFIRNSTQSMYCIP---CE-- 325
 QY 152 EWSFPWPKMKGKTCGFKGTRTETVRDILQHPSAKG-----KGNL 191
 DB 326 -----GPC---PRVCGDEBKTKTIDSVTAQMLQGCILKGNL 361
 RESULT 11
 FSPO XENLA
 ID FSPO XENLA STANDARD; PRT; 803 AA.
 AC P35447;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE F-spondin precursor.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=93376785; PubMed=8367492;
 RA Ruiz i Altaba A., Cox C., Jessell T.M., Klar A.;
 RT "Ectopic neural expression of a floor plate marker in frog embryos
 injected with the midline transcription factor Pintallavis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:8268-8272(1993)
 CC -!- FUNCTION: Promotes the attachment of spinal cord and sensory
 CC neuron cells and the outgrowth of neurites in vitro. May
 CC contribute to the growth and guidance of axons in both the spinal
 CC cord and the PNS.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: Expressed at high levels in the floor plate.
 CC -!- SIMILARITY: Contains 6 TSP type-1 domains.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC EMBL; L09123; AAA19105.1; -
 CC PIR; A47723; A47723.

DR InterPro; IPR002861; Reeler.
 DR InterPro; IPR000884; TSP1.
 DR Pfam; PF02014; Reeler; 1.
 DR Pfam; PF00090; TSP1; 6.
 DR SMART; SM00209; TSP1; 6.
 DR PROSITE; PS00092; TSP1; 6.
 KW Glycoprotein; Signal; Repeat; Cell adhesion.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 803 F-SPONDIN.
 FT DOMAIN 436 489 TSP TYPE-1 1.
 FT DOMAIN 495 549 TSP TYPE-1 2.
 FT DOMAIN 552 605 TSP TYPE-1 3.
 FT DOMAIN 608 662 TSP TYPE-1 4.
 FT DOMAIN 664 717 TSP TYPE-1 5.
 FT DOMAIN 750 802 TSP TYPE-1 6.
 FT CARBOHYD 210 210 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 677 677 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 803 AA; 90702 MM; D3A54E329548ED9 CRC64;
 Query Match 9.5%; Score 145.5; DB 1; Length 803;
 Best Local Similarity 26.6%; Pred. No. 0.0041;
 Matches 58; Conservative 23; Mismatches 84; Indels 53; Gaps 12;
 QY 57 CKPRLFFVLE-----RIGMKQIGVCLSSCPGYYGTR-YPDINKCTCKV---- 100
 DB 548 CEPSSCIVTEWAEWECSSATCRMGKKRHRMINTPAD--GSMCKADTTEVEKCMPECH 605
 QY 101 -----DCDTCFNNFCTCKSGFYHLGKCLDSCEGLEANN-----HTWBCV 143
 DB 606 TIPCVLSPWSEWSDCVTCGKTRTRQR---MLKSPSELGDCNEBELKQVEKCMLEPCP 662
 QY 144 SIHVCEASWSPWCPCKKCTCGFKGTRTETVRDILQHPSAKGKGNLCPTSETRACIV 203
 DB 663 --ISCLETWSYWSSEC---NKSG--KGHMIRMTMTWEPQFGA--VCPETVORKKC-- 711
 QY 204 QRKCKSGRGGKGRKRRKKLKNKRRKETSSTSSDSKG 241
 DB 712 RLKCKQK-----SSGNERHLKDKAREKSEKIKEDSDG 745
 RESULT 12
 NTC2 RAT
 ID NTC2 RAT STANDARD; PRT; 2471 AA.
 AC Q9QW10;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 2 precursor (Notch 2).
 GN NOTCH2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=93202015; PubMed=1295745;
 RA Weinmaster G., Roberts V.J., Lemke G.;
 RT "Notch2: a second mammalian Notch gene.";
 RL Development 116:931-941(1992).
 RN [2]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=21331789; PubMed=11438922;
 RA Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
 RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
 RT functional roles for the Notch-DSL signaling system during brain
 RT development.";
 RL J. Comp. Neurol. 436:167-181(2001).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.

CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs. May play an essential role in postimplantation
 CC development, probably in some aspect of cell specification and/or
 CC differentiation (By similarity).
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EGF) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -1- TISSUE SPECIFICITY: Highly expressed in the spleen and choroid
 CC plexus in the brain. Expressed in postnatal central nervous system
 CC (CNS) germinal zones and, in early postnatal life, within numerous
 CC cells throughout the CNS. It is more highly localized to
 CC ventricular germinal zones. Also found in the heart, liver and
 CC kidney.
 CC -1- DEVELOPMENTAL STAGE: Expressed in the brain during E14 and E17.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EGF). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -1- PTM: Phosphorylated (By similarity).
 CC -1- SIMILARITY: Belongs to the NOTCH family.
 CC -1- SIMILARITY: Contains 35 EGF-like domains.
 CC -1- SIMILARITY: Contains 2 Lin/Notch repeats.
 CC -1- SIMILARITY: Contains 6 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; M93661; AAK1358.1; --
 CC DR PIR; A49128; A49128.
 CC DR HSP; P00743; ICCP.
 CC DR InterPro; IPR002110; ANK.
 CC DR InterPro; IPR000152; Asx_hydroxyl_S.
 CC DR InterPro; IPR000742; EGF_2.
 CC DR InterPro; IPR001881; EGF_Ca.
 CC DR InterPro; IPR001438; EGF_II.
 CC DR InterPro; IPR006209; EGF_Like.
 CC DR InterPro; IPR002049; Laminin_EGF.
 CC DR InterPro; IPR008297; Notch.
 CC DR InterPro; IPR008000; Notch_dom.
 CC DR Pfam; PF00023; ank; 6.
 CC DR Pfam; PF00008; EGF; 35.
 CC DR Pfam; PF00066; notch; 2.
 CC DR PIRSF; PIRSF002279; Notch; 1.
 CC DR PRINTS; PR00010; EGFLOOD.
 CC DR PRINTS; PR00011; EGFAMININ.
 CC DR PRINTS; PR01452; NOTCH.
 CC DR SMART; SM00248; ANK; 6.
 CC DR SMART; SM00179; EGF_CA; 24.
 CC DR SMART; SM00004; NL; 2.
 CC DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 CC DR PROSITE; PS50088; ANK_REPEAT; 4.
 CC DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 CC DR PROSITE; PS00022; EGF_1; 34.
 CC DR PROSITE; PS01186; EGF_2; 26.
 CC DR PROSITE; PS50026; EGF_3; 35.
 CC DR PROSITE; PS01187; EGF_CA; 22.
 CC DR Receptor; Transcription regulation; Activator; Differentiation;
 CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
 CC Transmembrane; Glycoprotein; Signal; Phosphorylation.

FT	SIGNAL	1	25	POTENTIAL.
FT	CHAIN	26	2471	NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.
FT	CHAIN	1666	2471	NOTCH EXTRACELLULAR TRUNCATION (BY
FT				SIMILARITY).
FT	CHAIN	1697	2471	NOTCH INTRACELLULAR DOMAIN (BY
FT				SIMILARITY).
FT	DOMAIN	26	1677	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1678	1698	POTENTIAL.
FT	DOMAIN	1699	2471	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	26	63	EGF-LIKE 1.
FT	DOMAIN	64	102	EGF-LIKE 2.
FT	DOMAIN	105	143	EGF-LIKE 3.
FT	DOMAIN	144	180	EGF-LIKE 4.
FT	DOMAIN	182	219	EGF-LIKE 5.
FT	DOMAIN	221	258	EGF-LIKE 6.
FT	DOMAIN	260	296	EGF-LIKE 7.
FT	DOMAIN	298	336	EGF-LIKE 8.
FT	DOMAIN	338	374	EGF-LIKE 9.
FT	DOMAIN	375	413	EGF-LIKE 10.
FT	DOMAIN	415	454	EGF-LIKE 11.
FT	DOMAIN	456	492	EGF-LIKE 12.
FT	DOMAIN	494	530	EGF-LIKE 13.
FT	DOMAIN	532	568	EGF-LIKE 14.
FT	DOMAIN	570	605	EGF-LIKE 15.
FT	DOMAIN	607	643	EGF-LIKE 16.
FT	DOMAIN	645	680	EGF-LIKE 17.
FT	DOMAIN	682	718	EGF-LIKE 18.
FT	DOMAIN	720	755	EGF-LIKE 19.
FT	DOMAIN	757	793	EGF-LIKE 20.
FT	DOMAIN	795	831	EGF-LIKE 21.
FT	DOMAIN	833	871	EGF-LIKE 22.
FT	DOMAIN	873	909	EGF-LIKE 23.
FT	DOMAIN	911	947	EGF-LIKE 24.
FT	DOMAIN	949	985	EGF-LIKE 25.
FT	DOMAIN	987	1023	EGF-LIKE 26.
FT	DOMAIN	1025	1061	EGF-LIKE 27.
FT	DOMAIN	1063	1099	EGF-LIKE 28.
FT	DOMAIN	1101	1147	EGF-LIKE 29.
FT	DOMAIN	1149	1185	EGF-LIKE 30.
FT	DOMAIN	1187	1223	EGF-LIKE 31.
FT	DOMAIN	1225	1262	EGF-LIKE 32.
FT	DOMAIN	1264	1302	EGF-LIKE 33.
FT	DOMAIN	1304	1343	EGF-LIKE 34.
FT	DOMAIN	1374	1412	EGF-LIKE 35.
FT	DOMAIN	1645	1648	POLY-ALA.
FT	DOMAIN	1994	1997	POLY-LEU.
FT	DOMAIN	2426	2429	POLY-SER.
FT	DOMAIN	2446	2451	POLY-GLY.
FT	REPEAT	1420	1456	LIN/NOTCH 1.
FT	REPEAT	1503	1535	LIN/NOTCH 2.
FT	REPEAT	1827	1871	ANK 1.
FT	REPEAT	1876	1905	ANK 2.
FT	REPEAT	1909	1939	ANK 3.
FT	REPEAT	1943	1972	ANK 4.
FT	REPEAT	1976	2005	ANK 5.
FT	REPEAT	2009	2038	ANK 6.
FT	DISULFID	28	41	BY SIMILARITY.
FT	DISULFID	35	51	BY SIMILARITY.
FT	DISULFID	53	62	BY SIMILARITY.
FT	DISULFID	68	79	BY SIMILARITY.
FT	DISULFID	73	90	BY SIMILARITY.
FT	DISULFID	92	101	BY SIMILARITY.
FT	DISULFID	109	121	BY SIMILARITY.
FT	DISULFID	115	131	BY SIMILARITY.
FT	DISULFID	133	142	BY SIMILARITY.
FT	DISULFID	148	159	BY SIMILARITY.
FT	DISULFID	153	168	BY SIMILARITY.
FT	DISULFID	170	179	BY SIMILARITY.
FT	DISULFID	186	198	BY SIMILARITY.
FT	DISULFID	192	207	BY SIMILARITY.
FT	DISULFID	209	218	BY SIMILARITY.
FT	DISULFID	225	236	BY SIMILARITY.
FT	DISULFID	230	246	BY SIMILARITY.

```

FT DISULFID 248 257 BY SIMILARITY.
FT DISULFID 264 275 BY SIMILARITY.
FT DISULFID 269 284 BY SIMILARITY.
FT DISULFID 286 295 BY SIMILARITY.
FT DISULFID 302 315 BY SIMILARITY.
FT DISULFID 309 324 BY SIMILARITY.
FT DISULFID 326 335 BY SIMILARITY.
FT DISULFID 342 353 BY SIMILARITY.
FT DISULFID 347 362 BY SIMILARITY.
FT DISULFID 364 373 BY SIMILARITY.
FT DISULFID 379 390 BY SIMILARITY.
FT DISULFID 384 401 BY SIMILARITY.
FT DISULFID 403 412 BY SIMILARITY.
FT DISULFID 419 433 BY SIMILARITY.
FT DISULFID 427 442 BY SIMILARITY.
FT DISULFID 444 453 BY SIMILARITY.
FT DISULFID 460 471 BY SIMILARITY.
FT DISULFID 465 480 BY SIMILARITY.
FT DISULFID 482 491 BY SIMILARITY.
FT DISULFID 498 509 BY SIMILARITY.
FT DISULFID 503 518 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 536 547 BY SIMILARITY.

Query Match 9.4%; Score 144.5; DB 1; Length 2471;
Best Local Similarity 24.9%; Pred. No. 0.014;
Matches 55; Conservative 25; Mismatches 54; Indels 87; Gaps 14;

QY 30 QRRMHPNVSGGGGATCSDYNGCLSCRPFLPFVLERIGMKQIGVCLSCSPSGYGYTRY 89
DQ 947 QTDMECLSEPCNKG-GTCSDYVSYTC-----TCFAGPHGVHC 984
QY 90 P-DINKCTCKVDCTCFKNFC-----TKKSGFYHLGK 124
DQ 985 ENNIDECTE-----SSCFNGTCVDGINSFCLCPVGTGPPCLHHDINECSSNPLNSGT 1039
QY 125 CLD-----SCPEGLAANNHTMVCVSIH-CEASESWPSPCKMKGKTCGPKRGVTETRV 176
DQ 1040 CVDGLGTREYCTPLGYTGN-----CQTLVNLG-----SP-SPCKNGK-TCQAQKA----- 1083
QY 177 RDILOHPSAKGNLCPPTSETRTCTIVQRKCKSGKRGKGG 217
DQ 1084 -----RPRC-----LCPPGWDGAYCDVLNVSC-KAAALQNG 1113

RESULT 13
PSPO.RAT STANDARD; PRT; 807 AA.
AC P35446;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE F-spondin precursor.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryonic floor plate;
RX MEDLINE=92208952; PubMed=1555244;
RA Klar A., Baldassare M., Jessell T.M.;
RT "F-spondin: a gene expressed at high levels in the floor plate
RT encodes a secreted protein that promotes neural cell adhesion and
RT neurite extension."
RL Cell 69:95-110(1992).
CC -!- FUNCTION: Promotes the attachment of spinal cord and sensory
CC neuron cells and the outgrowth of neurites in vitro. May
CC contribute to the growth and guidance of axons in both the spinal
CC cord and the PNS.
CC -!- TISSUE SPECIFICITY: Expressed at high levels in the floor plate.
CC -!- SIMILARITY: Contains 6 TSP type-1 domains.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR M88469; AAA41174.1; -
DR PIR; A38152; A38152.
DR InterPro; IPR002861; Reeler.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF02014; Reeler; 1.
DR Pfam; PF00990; TSP1; 6.
DR SMART; SM00209; TSP1; 6.
DR PROSITE; PS0092; TSP1; 6.
KW Glycoprotein; Signal; Repeat; Cell adhesion.
FT SIGNAL 1 28
FT CHAIN 29 807 P-SPONDIN.
FT DOMAIN 442 455 TSP TYPE-1 1.
FT DOMAIN 501 555 TSP TYPE-1 2.
FT DOMAIN 558 611 TSP TYPE-1 3.
FT DOMAIN 614 666 TSP TYPE-1 4.
FT DOMAIN 668 721 TSP TYPE-1 5.
FT DOMAIN 754 806 TSP TYPE-1 6.
FT CARBOHYD 214 214 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 681 681 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 807 AA; 90773 MW; 3095259FAEAS9A CRC64;

Query Match 9.3%; Score 142.5; DB 1; Length 807;
Best Local Similarity 29.5%; Pred. No. 0.0066;
Matches 54; Conservative 16; Mismatches 60; Indels 53; Gaps 10;

QY 126 LDSCPEGLAANNHTM--ECVSIHVCSEASWSPCKMKGKTCGPKRGVTETRVRLDLOHP 183
DQ 647 LGDCNEDLQAEKMLPECP--IDCLSEWSQWSEC---NKSG--KGMIRTRTIQMEP 699
QY 184 SAKGKGLCPPTSETRTCTIVQRKCKSGK-----RGKKGRERKRLKKEB---- 229
DQ 700 QFGGAP--CPET-----VQRKCKRKLRLSPSIQKLWRREARRSRSEQLRHEBSDE 750
QY 230 -----RKETSSSDSKGLSSIEPTDQENKRRQOQKRRARD-----KQKSVSVS 276
DQ 751 QPFGCEMRPTAWSECTKLGCGGIQ-----ERYVTYKPKSKSQFTCKKKEIRAC 802
QY 277 TVH 279
DQ 803 NVH 805

RESULT 14
IGIR.MOUSE STANDARD; PRT; 1373 AA.
AC Q60751; O70438; Q62123;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Insulin-like growth factor I receptor precursor (EC 2.7.1.112).
GN IGIR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Navarro M., Garandel V., Barenton B., Bernardi H.;
RT "Cloning of cDNA for the mouse insulin-like growth factor I
RT receptor."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-329 FROM N.A.
RC STRAIN=CD-1; TISSUE=Kidney;

```

RA Jun W., Liu Z., Alvares K., Kumar A., Wallner B.J., Kanwar Y.S.;
 RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1134-1203 FROM N.A.
 RK MEDLINE=90152381; PubMed=2482828;
 RA Wilks A.F., Kurban R.R., Hovens C.M., Ralph S.J.;
 RT "The application of the polymerase chain reaction to cloning members
 of the protein tyrosine kinase family.";
 RL Gene 85:67-74(1989).
 CC -1- FUNCTION: THIS RECEPTOR BINDS INSULIN-LIKE GROWTH FACTOR I (IGF I)
 WITH A HIGH AFFINITY AND IGF II WITH A LOWER AFFINITY. IT HAS A
 TYROSINE-PROTEIN KINASE ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 tyrosine phosphate.
 CC -1- SUBUNIT: Tetramer of 2 alpha and 2 beta chains linked by disulfide
 bonds. The alpha chains contribute to the formation of the ligand-
 binding domain, while the beta chain carries the kinase domain.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: Belongs to the Tyr family of protein kinases. Insulin
 receptor subfamily.
 CC -1- SIMILARITY: Contains 2 fibronectin type III domains.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF056187; AAC12782.1; -.
 DR EMBL; U00182; AAC52123.1; -.
 DR EMBL; M33422; AAA40013.1; -.
 DR PIR; A48805; A48805.
 DR HSP; P06213; IIRK.
 DR MGD; MGI:96433; Igflr.
 DR GO; GO:0009887; P: organogenesis; IMP.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR008957; FN III-like.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR006211; Furin-like.
 DR InterPro; IPR006212; Furin repeat.
 DR InterPro; IPR009030; Grow_fac_recep.
 DR InterPro; IPR000719; Prot_kinase.
 DR InterPro; IPR002011; Recepttyr_kinII.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR InterPro; IPR008266; Tyr_pkinase_AS.
 DR Pfam; PF00041; fn3; 2.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF01030; Recep_L domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR SMART; SM00060; FN3; 3.
 DR SMART; SM00261; FU; 1.
 DR SMART; SM00219; TyrKc; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
 DR PROSITE; PS00111; PROTEIN KINASE_DOM; 1.
 DR PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
 DR Transferrase; Tyrosine-protein kinase; Receptor; Transmembrane;
 Glycoprotein; ATP-binding; phosphorylation; Repeat; Signal.
 KW SIGNAL 1 30
 FT CHAIN 31 741 INSULIN-LIKE GROWTH FACTOR I RECEPTOR,
 FT CHAIN 742 1373 ALPHA-CHAIN
 FT CHAIN 742 936 INSULIN-LIKE GROWTH FACTOR I RECEPTOR,
 FT DOMAIN 742 936 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 937 960 BETA-CHAIN.
 FT DOMAIN 961 1373 CYTOPLASMIC (BY SIMILARITY).
 FT DOMAIN 608 829 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 830 929 FIBRONECTIN TYPE-III 2.
 FT DOMAIN 1000 1276 PROTEIN KINASE.

FT NP_BIND 1006 1014 ATP (BY SIMILARITY).
 FT BINDING 1034 1034 ATP (BY SIMILARITY).
 FT ACT_SITE 1137 1137 BY SIMILARITY.
 FT DISULFID 215 224 BY SIMILARITY.
 FT DISULFID 219 230 BY SIMILARITY.
 FT DISULFID 231 239 BY SIMILARITY.
 FT DISULFID 235 248 BY SIMILARITY.
 FT DISULFID 251 260 BY SIMILARITY.
 FT DISULFID 264 276 BY SIMILARITY.
 FT DISULFID 282 303 BY SIMILARITY.
 FT DISULFID 307 321 BY SIMILARITY.
 FT DISULFID 324 328 BY SIMILARITY.
 FT CARBOHYD 51 51 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 102 102 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 135 135 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 245 245 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 314 314 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 439 439 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 535 535 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 608 608 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 641 641 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 748 748 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 757 757 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 765 765 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 901 901 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 914 914 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT MOD_RES 1167 1167 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CONFLICT 58 59 FL -> LV (IN REF. 2).
 FT CONFLICT 260 260 C -> S (IN REF. 2).
 FT CONFLICT 301 301 D -> G (IN REF. 2).
 FT CONFLICT 306 306 E -> V (IN REF. 2).
 FT CONFLICT 324 324 C -> S (IN REF. 2).
 FT CONFLICT 1134 1134 V -> I (IN REF. 3).
 FT CONFLICT 1145 1145 V -> D (IN REF. 3).
 FT CONFLICT 1202 1202 V -> I (IN REF. 3).
 SQ SEQUENCE 1373 AA; 155787 MW; 58E3B72EF10E379 CRC64;
 Query Match 9.3%; Score 142; DB 1; Length 1373;
 Best Local Similarity 28.0%; Pred. No. 0.012;
 Matches 46; Conservative 20; Mismatches 58; Indels 40; Gaps 11;
 Qy 34 HPNVSGGCGCATCSYNGCLSCFPLFFVLEIGMKQIGVCLSSCPSSGYGTYPDIN 93
 Db 232 HPE-----CLGSCHTPDNTTCVACRHHY-----KGVCPACPGTY--RF----- 271
 Qy 94 KCTCKVDCTCFN-KNFTCKSGFYHLHGKLDSCPEGLEANN-HTMBCVSVHVCBAS 151
 Db 272 EGWRC-VDRDFCANIPNABSSDSOGFVHDECHQBCPSGFIKSTQSMYCIP---CB--- 325
 Qy 152 EWSWSPQMKKKGKTCGFKRGTTETVRDILQHPKAG-----KGNL 191
 Db 326 -----GFC---PKVCGDEBKTKTIDSVTSQMLQGCTILKGNL 361
 RESULT 15
 NTC2 MOUSE
 ID NTC2 MOUSE STANDARD; PRT; 2470 AA.
 AC O35516; Q06008; Q60941.
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DB Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Motch
 B).
 GN NOTCH2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CS7BL/6; TISSUE=Thymus;

RA Hamada Y., Higuchi M., Tsujimoto Y.;
RT "Complete amino acid sequence and multiform transcripts encoded by a
RT single copy of mouse Notch2 gene.";
RN Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [2].
RC SEQUENCE OF 316-1518 FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
RX MEDLINE=93178563; PubMed=8440332;
RA Lardelli M., Lendahl U.;
RT "Moth and Notch B-two mouse Notch homologues coexpressed in a
RT wide variety of tissues.";
RL Exp. Cell Res. 204:364-372(1993).
RN [3].
RN SEQUENCE OF 1765-2153 FROM N.A.
RX MEDLINE=97075110; PubMed=8917536;
RA Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,
RA Martin D.I.;
RT "Inhibition of granulocytic differentiation by mNotch1.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).
RN [4].
RN FUNCTION.
RX MEDLINE=99396706; PubMed=10393120;
RA Hamada Y., Kadokawa Y., Okabe M., Ikawa M., Coleman J.R.,
RA Tsujimoto Y.;
RT "Mutation in ankyrin repeats of the mouse Notch2 gene induces early
RT embryonic lethality.";
RL Development 126:3415-3424 (1999).
RN [5].
RN DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.
RX MEDLINE=95333893; PubMed=7609614;
RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;
RT "Differential expression of Notch1 and Notch2 in developing and adult
RT mouse brain.";
RL Brain Res. Mol. Brain Res. 29:263-272(1995).
RN [6].
RN POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
RX MEDLINE=21523956; PubMed=11518718;
RA Saxena M.T., Schroter E.H., Mumm J.S., Kopan R.;
RT "Murine notch homologs (NL-4) undergo presenilin-dependent
RT proteolysis.";
RL J. Biol. Chem. 276:40268-40273(2001).
RN [7].
RN POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
RX MEDLINE=21374376; PubMed=11459941;
RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
RT "Conservation of the biochemical mechanisms of signal transduction
RT among mammalian Notch family members.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity). May play an essential role in
CC postimplantation development, probably in some aspect of cell
CC specification and/or differentiation.
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC terminal fragment N(EC) which are probably linked by disulfide
CC bonds.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=035516-1; Sequence=Displayed;
CC Name=2;
CC IsoId=035516-2; Sequence=VSP_001405;
CC Note=No experimental confirmation available;
CC -1- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,
CC neuroepithelia, somites, optic vesicles and branchial arches, but
CC not heart.
CC -1- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,

CC the postnatal ependymal cells, and the choroid plexus throughout
CC embryonic and postnatal development.
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC ligand binding, it is cleaved by TMP-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane.
CC -1- PTM: Phosphorylated.
CC -1- SIMILARITY: Belongs to the NOTCH family.
CC -1- SIMILARITY: Contains 35 EGF-like domains.
CC -1- SIMILARITY: Contains 2 Lin/Notch repeats.
CC -1- SIMILARITY: Contains 6 ANK repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D32210; BAA22094.1; -;
DR EMBL; X68279; CAA48340.1; -;
DR EMBL; U31881; AAC52924.1; -;
DR PIR; A49175; A49175.
DR HSP; P16109; IFSB.
DR MGD; MGI:97364; Notch2.
DR GO; GO:0005887; C:integral to plasma membrane; IC.
DR GO; GO:0005155; P:protein binding; IPI.
DR GO; GO:0002011; P:morphogenesis of an epithelial sheet; IMP.
DR GO; GO:0007219; P:N signaling pathway; IC.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; ANK_hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR008297; Notch.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00066; notch; 2.
DR PIRSF; PIRSF002279; Notch; 1.
DR PRINTS; PRO0010; EGFLOOD.
DR PRINTS; PRO0011; EGF_LAMININ.
DR PRINTS; PRO1452; NOTCH.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00179; EGF_CA; 23.
DR SMART; SM00004; NL_3.
DR PROSITE; PS0297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00088; ANK_REPEAT; 4.
DR PROSITE; PS00010; ASX_HYDROXYL; 22.
DR PROSITE; PS00022; EGF_1; 33.
DR PROSITE; PS01186; EGF_2; 27.
DR PROSITE; PS00026; EGF_3; 35.
DR PROSITE; PS01187; EGF_CA; 22.
KW Receptor; Transcription regulation; Activator; Differentiation;
KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Glycoprotein; Signal; Phosphorylation;
KW Alternative splicing.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 2470 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.
FT CHAIN 1666 2470 NOTCH EXTRACELLULAR TRUNCATION.
FT CHAIN 1697 2470 NOTCH INTRACELLULAR DOMAIN.
FT DOMAIN 26 1677 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1678 1698 POTENTIAL.

FT	DOMAIN	1699	2470	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	26	63	EGF-LIKE 1.
FT	DOMAIN	64	102	EGF-LIKE 2.
FT	DOMAIN	105	143	EGF-LIKE 3.
FT	DOMAIN	144	180	EGF-LIKE 4.
FT	DOMAIN	182	219	EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	221	256	EGF-LIKE 6 (INCOMPLETE).
FT	DOMAIN	258	294	EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	296	334	EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	336	372	EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	373	411	EGF-LIKE 10.
FT	DOMAIN	413	452	EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	454	490	EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	492	528	EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	530	566	EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	568	603	EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	605	641	EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	643	678	EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	680	716	EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	718	753	EGF-LIKE 19.
FT	DOMAIN	755	791	EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	793	829	EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	831	869	EGF-LIKE 22.
FT	DOMAIN	871	907	EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	909	945	EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	947	983	EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	985	1021	EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1023	1059	EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1061	1097	EGF-LIKE 28.
FT	DOMAIN	1099	1145	EGF-LIKE 29.
FT	DOMAIN	1147	1183	EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1185	1221	EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1223	1260	EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1262	1300	EGF-LIKE 33.
FT	DOMAIN	1302	1345	EGF-LIKE 34.
FT	DOMAIN	1372	1410	EGF-LIKE 35.
FT	REPEAT	1418	1454	LIN/NOTCH 1.
FT	REPEAT	1501	1533	LIN/NOTCH 2.
FT	REPEAT	1825	1869	ANK 1.

Query Match 9.2%; Score 140.5; DB 1; Length 2470;

Best Local Similarity 24.4%; Pred. No. 0.026;

Matches 54; Conservative 24; Mismatches 56; Indels 87; Gaps 14;

QY	30	QRRMHPNVSGCGGCGCATCDYNGCLSCPKRLPFVLERIGHKQIGVCLSSCPSSGYGTRY	89
Db	945	QTDNNECLSEPCCKNG-GTCSDYNSVTC-----TCPAGFHGVHC	982
QY	90	P-DINKCTKVCDCDTCFNKFC-----TKCKSGFYHLGK	124
Db	983	ENNIDECTE-----SSCFNGGTCVDGINSFSCICPVGFTGPFCLHDINECSNPLNAGT	1037
QY	125	CLDS-----CPGLEANNHMECVSIVH-CEASESWSPWSPCKKGTGFKRGTTETRV	176
Db	1038	CVDGLGYRCICPLGYTGKY-----CQTLVNLCSR-----SPCKNG-TC-----VQEKA	1081
QY	177	RDILQHPASAKGNLCPPTSETETCTIVQRKCKSKGERKKG	217
Db	1082	R-----PHC-----LCPPGWDGAYCDVNLVSC-KAALQKG	1111

Search completed: June 29, 2004, 17:00:17
Job time : 9.46481 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 16:55:24 ; Search time 35.5522 Seconds
(without alignments)
2476.067 Million cell updates/sec

Title: US-09-894-912A-32

Perfect score: 1535

Sequence: 1 MHLRLISCFIILNFMFYIG.....QQRRARDKQKSVSVSTVH 279

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1315.5	85.7	272	4 Q9BX4	Q9bx4 homo sapien
2	1273	82.9	292	4 Q96K87	Q96k87 homo sapien
3	1212	79.0	217	11 Q9CSB2	Q9csb2 mus musculus
4	1107	72.1	224	11 Q8BVW2	Q8bv2 mus musculus
5	632	41.2	265	11 Q9Z132	Q9z132 mus musculus
6	616.5	40.2	236	4 Q8N7L5	Q8n715 homo sapien
7	565.5	36.8	243	11 Q8BFU0	Q8bfu0 mus musculus
8	450	29.3	224	4 Q9UGB2	Q9ugb2 homo sapien
9	449.5	29.3	176	4 Q8N6X6	Q8n6x6 homo sapien
10	419.5	27.3	152	11 Q7TPX3	Q7tpx3 mus musculus
11	283.5	18.5	138	11 Q8BU73	Q8bu73 mus musculus
12	185.5	12.1	1299	5 Q26489	Q26489 spidoptera
13	174	11.3	1376	5 Q8S252	Q8s252 spidoptera
14	171	11.1	296	11 Q35171	Q35171 mus musculus
15	171	11.1	826	11 Q8CF22	Q8cf22 mus musculus
16	171	11.1	932	11 Q62030	Q62030 mus musculus

17	165	10.8	915	11	Q91VK0	Q91vk0 mus musculus
18	165.5	10.8	913	13	Q8AY18	Q8ay18 rana esculle
19	162.5	10.6	1101	5	Q964D2	Q964d2 entamoeba h
20	161.5	10.5	803	13	Q42114	Q42114 brachydanio
21	160.5	10.5	503	5	Q9U018	Q9u018 giardia lam
22	160.5	10.5	1074	5	Q964D1	Q964d1 entamoeba h
23	160	10.4	808	13	Q42113	Q42113 brachydanio
24	154	10.0	913	4	Q968P4	Q968p4 homo sapien
25	154	10.0	4010	11	Q80T14	Q80t14 mus musculus
26	151.5	9.9	3869	5	Q86PQ3	Q86pq3 cryptospori
27	149	9.7	1371	11	Q9QVW4	Q9qv4 rattus sp.
28	147.5	9.6	548	5	Q9GQ45	Q9gq45 giardia lam
29	144	9.4	1362	13	Q9PVZ4	Q9pv24 xenopus lae
30	143	9.3	4007	4	Q86XX4	Q86xx4 homo sapien
31	142	9.3	898	5	Q76822	Q76822 branchiosco
32	141.5	9.2	402	11	Q8K2Q8	Q8k2q8 mus musculus
33	141.5	9.2	807	11	Q8VCC9	Q8vcc9 mus musculus
34	140.5	9.2	213	11	Q991E4	Q991e4 mus musculus
35	140.5	9.2	214	11	Q8R0Y1	Q8r0y1 mus musculus
36	140.5	9.2	660	5	Q23832	Q23832 cryptospori
37	139.5	9.1	807	4	Q9HCB6	Q9hcb6 homo sapien
38	139	9.1	1193	5	Q9Y1X8	Q9y1x8 ephydatia f
39	138.5	9.0	724	4	Q94862	Q94862 homo sapien
40	138.5	9.0	807	4	Q8NCD7	Q8ncd7 homo sapien
41	137	8.9	440	5	O18003	O18003 caenorhabdi
42	137	8.9	441	4	Q96JW7	Q96jw7 homo sapien
43	137	8.9	651	4	Q86UZ4	Q86uz4 homo sapien
44	136.5	8.9	642	5	Q8MPM6	Q8mpm6 giardia lam
45	136.5	8.9	739	5	Q9G824	Q9g824 giardia lam

ALIGNMENTS

RESULT 1

Q9BX4 PRELIMINARY; PRT; 272 AA.
AC Q9BX4, 2001 (TREMREL. 17, Created)
DT 01-JUN-2001 (TREMREL. 17, Last sequence update)
DT 01-OCT-2003 (TREMREL. 25, Last annotation update)
DB Thrombospondin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Mao Y., Xie Y., Zhou Z., Zhao W., Wang S., Huang Y., Wang S.,
Tang R., Chen X., Wu C.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF251057; AAH22367.1; -.
DR EMBL; BC022367; AAH22367.1; -.
DR Genew; HGNC:20866; THSD2.
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR InterPro; IPR00884; TSP1.
DR Pfam; PF00090; tsp_1; 1.
DR SMART; SM00261; FU; 2.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS50092; TSP1; 1.
SQ SEQUENCE 272 AA; 30928 MW; CACAC6B7E781189 CRC64;

Query Match 85.7%; Score 1315.5; DB 4; Length 272;

Best Local Similarity 87.1%; Pred. No. 1.8e-110;

Matches 243; Conservative 11; Mismatches 18; Indels 7; Gaps 4;

QY 1 MHLRLISCFIILNFMFYIGSQRRARDKQKSVSVSTVHNGCLSKPR 60


```

Db      1  MLRLISWFIILNFMEYIGSNASGRQRMRHPNVSGCGGCATCDYNGCLSCKPR 60
QY      61  LFPVLERIGMKQIGVCLSSPCSGYGYTRYDINKTKCKVDCTCNKFNCTKCKSGFYL 120
Db      61  LFPALERIGMKQIGVCLSSPCSGYGYTRYDINKTKCKADCTCNKFNCTKCKSGFYL 120
QY      121  HLGKCLDSCPEGLANNTMBCVSIHVCEASEWSPMSCKKGTCTGKRGTTETRVRDIL 180
Db      121  HLGKCLDNCPEGLANNTMBCVSIHVCESEWNPSPCTKGTCTGKRGTTETRVREII 180
QY      181  QHPKAGKGNLCPPTSETTCTIVQRKCKSGRGGKGRKRKLNKBERKETSSSDSK 240
Db      181  QHPSA--KGNLCPPTNETRKTIVQRKCKSGRGGKGRKRKLNKBERKETSSSDSK 240
QY      241  GLESSIETPDQENKEROQKQKRDARKKOKSVSVSTVH 279
Db      237  SLESSKEIPEQRENK--QOKKRVQDK-QKSVSVSTVH 272

```

RESULT 2

```

Q96K87 PRELIMINARY; PRT; 292 AA.
AC Q96K87;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ14440.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y., Oshima A.;
RT "NEO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027346; BAB55051.1; -.
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR InterPro; IPR000884; TSPI.
DR Pfam; PF00090; tsp_1; 1.
DR SMART; SM00261; FU; 2.
DR SMART; SM00209; TSPI; 1.
DR PROSITE; PS50092; TSPI; 1.
KW Hypothetical protein.
SQ SEQUENCE 292 AA; 33243 MW; 01E2774AC3D4A6F8 CRC64;

```

Query Match 82.4%; Score 1273; DB 4; Length 292;
 Best Local Similarity 84.4%; Pred. No. 1.3e-106;
 Matches 233; Conservative 14; Mismatches 23; Indels 6; Gaps 3;

```

QY      1  MLRLISCFPIILNFMEYIGSNASGRQRMRHPNVSGCGGCATCDYNGCLSCKPR 60
Db      1  MLRLISWFIILNFMEYIGSNASGRQRMRHPNVSGCGGCATCDYNGCLSCKPR 60
QY      61  LFPVLERIGMKQIGVCLSSPCSGYGYTRYDINKTKCKVDCTCNKFNCTKCKSGFYL 120
Db      61  LFPALERIGMKQIGVCLSSPCSGYGYTRYDINKTKCKADCTCNKFNCTKCKSGFYL 120
QY      121  HLGKCLDSCPEGLANNTMBCVSIHVCEASEWSPMSCKKGTCTGKRGTTETRVRDIL 180
Db      121  HLGKCLDNCPEGLANNTMBCVSIHVCESEWNPSPCTKGTCTGKRGTTETRVREII 180
QY      181  QHPKAGKGNLCPPTSETTCTIVQRKCKSGRGGKGRKRKLNKBERKETSSSDSK 240
Db      181  QHPSA--KGNLCPPTNETRKTIVQRKCKSGRGGKGRKRKLNKBERKETSSSDSK 240

```

```

QY      241  GLESSIETPDQENKEROQKQKRDARKKOKSVSVS 276
Db      237  SLESSKEIPEQRENK--QOKKRVQDKQKSGIEVT 270

```

RESULT 3

```

Q9CSB2 PRELIMINARY; PRT; 217 AA.
AC Q9CSB2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE THSD2 OR 2810459H04RIK.
GN Mus musculus (Mouse)
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Mameda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleschmann W., Gaasterland T., Gibbs C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK013366; BAB28811.1; -.
DR MGD; MGI:1920030; Thsd2.
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR InterPro; IPR000884; TSPI.
DR Pfam; PF00090; tsp_1; 1.
DR SMART; SM00261; FU; 2.
DR SMART; SM00209; TSPI; 1.
DR PROSITE; PS50092; TSPI; 1.
FT NON_TER 217 217
SQ SEQUENCE 217 AA; 24304 MW; 0DCF938E9B3BF7 CRC64;

```

Query Match 79.0%; Score 1212; DB 11; Length 217;
 Best Local Similarity 98.6%; Pred. No. 3e-101;
 Matches 216; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

```

QY      1  MLRLISCFPIILNFMEYIGSNASGRQRMRHPNVSGCGGCATCDYNGCLSCKPR 60
Db      1  MLRLISWFIILNFMEYIGSNASGRQRMRHPNVSGCGGCATCDYNGCLSCKPR 60
QY      61  LFPVLERIGMKQIGVCLSSPCSGYGYTRYDINKTKCKVDCTCNKFNCTKCKSGFYL 120
Db      61  LFPALERIGMKQIGVCLSSPCSGYGYTRYDINKTKCKADCTCNKFNCTKCKSGFYL 120
QY      121  HLGKCLDSCPEGLANNTMBCVSIHVCEASEWSPMSCKKGTCTGKRGTTETRVRDIL 180
Db      121  HLGKCLDNCPEGLANNTMBCVSIHVCESEWNPSPCMKGTCTGKRGTTETRVRDIL 180
QY      181  QHPKAGKGNLCPPTSETTCTIVQRKCKSGRGGKGRKRKLNKBERKETSSSDSK 219
Db      181  QHPSA--KGNLCPPTSETTCTIVQRKCKSGRGGKGRKRKLNKBERKETSSSDSK 217

```

RESULT 4

Q8BVW2 PRELIMINARY; PRT; 224 AA.
 AC Q8BVW2; TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Thrombospondin homolog.
 GN THSD2 OR 2810459H04RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 [1]
 SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Skin;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573(2002).
 DR EMBL; AK076308; BAC36296.1; -.
 DR MGD; MGI:1920030; Thsd2.
 DR InterPro; IPR006212; Furin repeat.
 DR InterPro; IPR009030; Grow_fac_recep.
 DR SMART; SM00261; FU; 2.
 DR SEQUENCE 224 AA; 25398 MW; BCL3B083497CFEE3 CRC64;

Query Match 72.1%; Score 1107; DB 11; Length 224;
 Best Local Similarity 88.4%; Pred. No. 8.7e-92;
 Matches 198; Conservative 5; Mismatches 19; Indels 2; Gaps 1;

Qy 1 MHLRLISCFPIILNFMFYIGSNASRGRRORRMPHNPVNSQGGCGATCSDYNGCLSKPRL 60
 Db 1 MHLRLISCFPIILNFMFYIGSNASRGRRORRMPHNPVNSQGGCGATCSDYNGCLSKPRL 60
 Qy 61 LFFVLERIGMKQIGVCLSSCPGSGYGYTRYDINKCTKVDCTCFNKNFCTCKSGPYL 120
 Db 61 LFFVLERIGMKQIGVCLSSCPGSGYGYTRYDINKCTKVDCTCFNKNFCTCKSGPYL 120
 Qy 121 HLKCLDSCPGLRANNHMCVSIHVCEASESPWSPCMKKGKTCGFKGTETVRDIL 180
 Db 121 HLKCLDSCPGLRANNHMCVSIHVCEASESPWSPCMKKGKTCGFKGTETVRDIL 180
 Qy 181 QHPSAKGKGNLCPTSETRTCTIVQKCKSGRGGKGRKRRKXK 224
 Db 181 QHPSA--KGNPVPNQDRDNLYSTKGVFKGRARKGKRETKK 222

RESULT 5

Q92132 PRELIMINARY; PRT; 265 AA.
 AC Q92132;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Thrombospondin type 1 domain.
 GN RSPONDIN OR R-SPONDIN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 [1]
 SEQUENCE FROM N.A.
 RA Kanata T., Katsube K., Michikawa M., Yamada M., Mizusawa H.;
 RP "R-spondin, a novel thrombospondin type 1 domain gene, expressed in
 RT the dorsal neural tube";
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB016768; BAA75640.1; -.
 DR MGD; MGI:2183426; Repondin.

DR InterPro; IPR006212; Furin repeat.
 DR InterPro; IPR009030; Grow_fac_recep.
 DR SMART; SM00261; FU; 1.
 DR SMART; SM00261; FU; 2.
 DR SMART; SM00209; TSP1; 1.
 DR PROSITE; PS50092; TSP1; 1.
 DR SEQUENCE 265 AA; 29331 MW; PFEB8964743F5963 CRC64;
 Query Match 41.2%; Score 632; DB 11; Length 265;
 Best Local Similarity 43.7%; Pred. No. 5.8e-49;
 Matches 114; Conservative 42; Mismatches 91; Indels 14; Gaps 4;

Qy 3 LRLISCFPIILNFMFYIGSNASRGRRORRMPHNPVNSQGGCGATCSDYNGCLSKPRL 61
 Db 1 MRLGLCVVALVLSWTHIAVSGRIGKGRRIASBGSQACAKGCLCSVNGCLCSPKL 60
 Qy 62 FFVLERIGMKQIGVCLSSCPGSGYGYTRYDINKCTKVDCTCFNKNFCTCKSGPYL 120
 Db 61 FILLERNDIRGVCLPSCPFGYDARNPDMNKKICKLIEHCEAFSHNFTCKQBALYL 120
 Qy 121 HLKCLDSCPGLRANNHMCVSIHVCEASESPWSPCMKKGKTCGFKGTETVRDIL 180
 Db 121 HKGRCPACPEGSTAANSTMEGSPACQCESESPWSPCMKKGKTCGFKGTETVRDIL 180
 Qy 181 QHPSAKGKGNLCPTSETRTCTIVQKCKSGRGGKGRKRRKXK 240
 Db 181 HAPG--GDHTTCSDTKTRCTVTRTPCPGQKRGKGGQGRRENANRHPARKNSKEPRS- 237
 Qy 241 GLESSIETPDQENKEROQQ 261
 Db 238 -----NSRRHKGQOOPQ 249

RESULT 6

Q8N7L5 PRELIMINARY; PRT; 236 AA.
 AC Q8N7L5;
 DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Hypothetical protein FLJ40906.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A.
 RC TISSUE=Uterus;
 RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
 RA Arita M., Mueashino K., Yuuki H., Hara H., Sugiyama T., Irie K.,
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
 RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
 RA Sugano S., Nagahori K., Masuho Y., Nagai K., Isegai T.;
 RL "NEDO human cDNA sequencing project";
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK098225; BAC05263.1; -.
 DR InterPro; IPR006212; Furin repeat.
 DR InterPro; IPR009030; Grow_fac_recep.
 DR InterPro; IPR009030; Grow_fac_recep.
 DR Pfam; PF00090; tsp_1; 1.
 DR SMART; SM00261; FU; 2.
 DR SMART; SM00209; TSP1; 1.
 DR PROSITE; PS50092; TSP1; 1.
 DR Hypothetical protein.
 KW SEQUENCE 236 AA; 25942 MW; 8D03803127EC5678 CRC64;
 Query Match 40.2%; Score 616.5; DB 4; Length 236;
 Best Local Similarity 47.3%; Pred. No. 1.3e-47;
 Matches 112; Conservative 32; Mismatches 72; Indels 21; Gaps 4;

QY 32 RHHPNVSCQGGCATCSYNGCLSCPKPLFFVLERIGMKQIGVCLSSCPGSGYGYTRPD 91
 Db 4 RVSAGSQACAKGCLSEVNGCLSKSPKFIILLERNDIRQVGVCLSCPPGPFDAENPD 63
 QY 92 INKTKCKVD-CDTCPNNKFTCKSGFYHLGKCLDSCPEGLNHNTHMTCVSIHVCEA 150
 Db 64 MNKCIKIEHCACPFHNFCTKCEGLYHLKGRYCPACPEGSSAANGTSCSSPAQCEV 123
 QY 151 SEWSPWSPCMKKGKTCFPGKGTETVRDILQHPGAKGNLCPTSTETRTCTIVORCKSK 210
 Db 124 SEWSPWSPCMKKGKTCFPGKGTETVRDILQHPGAKGNLCPTSTETRTCTIVORCKSK 210
 QY 211 GEGKKG---RRKRKLANKERKTSSSSDSGLESSIETTPDQGNKEROQOQKR 263
 Db 182 GQYKGGGQGERNANRNLRKESKEAGCS-----RRKKGQOQOQOQ 224

RESULT 7

Q8BFUO PRELIMINARY; PRT; 243 AA.
 AC Q8BFUO
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Hypothetical thrombospondin type I repeat.
 GN 2610028P08RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Eye, and Hippocampus;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK049891; BAC33974.1; -
 DR EMBL; AK087485; BAC39893.1; -
 DR MGD; MGI:1922667; 2610028P08RIK.
 DR InterPro; IPR006212; Furin repeat.
 DR InterPro; IPR009030; Grow_fac_recep.
 DR InterPro; IPR00884; TSPI.
 DR SMART; SM00261; FU; 2.
 DR SMART; SM00209; TSPI; 1.
 DR PROSITE; PS50092; TSPI; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 243 AA; 28275 MW; ED76A08D61012ED7 CRC64;

Query Match 36.8%; Score 565.5; DB 11; Length 243;
 Best Local Similarity 45.8%; Pred. No. 5.2e-43;
 Matches 110; Conservative 36; Mismatches 83; Indels 11; Gaps 7;
 QY 1 MLRLISCFPIILNFMVEYIGSQNAGRRORRHHPNVSCQGGCATCSYNGCLSCPKR 60
 Db 1 MRCLSPFALLIINCNDYSQC-GNRRNRKA-SYVSNPICKGLSCSDKDGSCRCQK 58
 QY 61 LFFVLERIGMKQIGVCLSSCPGSGYGYTRYDINKTKCKVD-CDTCPNNKFTCKSGFY 119
 Db 59 LFFFLAREGWQYGECLHSCPSGYGYGHRAFDNMRACRIENCSDSCFSDCTCKVGFY 118
 QY 120 LHLGKCLDSCPEGLNHNTHMTCVSIHVCEASWSPWSPCMKKGKTCFPGKGTETVRDI 179
 Db 119 LHRGRCFDECPDGFAPLDETMCVB--GCEVGHSEWGTCSRNRNRCGPKWGLETRQI 176
 QY 180 LQHPGAKGNL-CPTSTETRTCTIVORCKSKGERGKGRERKRLNKE--ERKETSSS 236
 Db 177 VKAPA---KDTICPTIATESRCMMNRHCPGGKRTPKAKEKRNKGRKRLIERAQEHS 233
 RESULT 8

Q9UGB2 PRELIMINARY; PRT; 224 AA.
 AC Q9UGB2
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE D7824P16.3 (Novel protein similar to mouse thrombospondin type 1
 DE domain protein R-spondin) (Fragment).
 GN D7824P16.3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Blakey S.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL050325; CAB65783.3; -
 DR Genew; HGNC:16175; C20orf182.
 DR InterPro; IPR006212; Furin repeat.
 DR InterPro; IPR009030; Grow_fac_recep.
 DR InterPro; IPR00884; TSPI.
 DR SMART; SM00261; FU; 2.
 DR SMART; SM00209; TSPI; 1.
 DR PROSITE; PS50092; TSPI; 1.
 FT NON TER 224 224
 SQ SEQUENCE 224 AA; 25042 MW; 97D26AD34CDBF812 CRC64;
 Query Match 29.3%; Score 450; DB 4; Length 224;
 Best Local Similarity 42.1%; Pred. No. 1.2e-32;
 Matches 88; Conservative 33; Mismatches 70; Indels 18; Gaps 7;
 QY 28 RRQRHHPNVSCQGGCATCSYNGCLSCPKPLFFVLERIGMKQIGVCLSSCPGSGYGT 87
 Db 22 RRKXQVGTGLGNCCT-GCIICSENGCSTCOQLFLFIRREGIKQYKGLHDCPPGYGI 80
 QY 88 RYDPINKTKCKVD-CDTCPNNKFTCKSGFYHLGKCLDSCPEGLNHNTHMTCVSIHV 147
 Db 81 RGQEVNRCKKGATCSGSCFSDFCIRCKRQFYLYKGLCTCPPTGLAHQNTRECOG--B 138
 QY 148 CEASWSPWSPCMKKGKTCFPGKGTETVRDI--LQHPGAKGNLCPTSTETRTCTIVOR 205
 Db 139 CELPGWGSPTCHNGKTCGSAWGLSESRVREAGRACHEAA----TCQVLSESRKCPQR 194
 QY 206 KCKSKGERG---KKGRE----RKRKLNK 227
 Db 195 P--CPGERSPGQKGRKRRPRKDKLDR 221
 RESULT 9
 Q8NEX6 PRELIMINARY; PRT; 176 AA.
 AC Q8NEX6
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Similar to putative.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Strausberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC027938; AAT27938.1; -
 DR InterPro; IPR006212; Furin repeat.
 DR InterPro; IPR009030; Grow_fac_recep.
 DR InterPro; IPR00884; TSPI.
 DR SMART; SM00261; FU; 1.
 DR SMART; SM00209; TSPI; 1.
 DR PROSITE; PS50092; TSPI; 1.

[illegible]

[illegible]

155 FDSELVGCGCHHTCRTCVGSRRBCHCAKSFHFQDMKVCAPAGCSGFTPEEMPGLPHKV 224
159 CMKKGKTC---GFKRG-----TETVRVDILQHPSAKGNLCPPTSETATCTCIQVRK 206
225 GRCBENCILCEGSSRNCRCAGFTOLGTSTCIINHCTCSNADETFCEWVKSNRLC--ERK 282

RESULT	ID	Q8CF22	PRELIMINARY:	PRT:	826 AA.
Q8CF22	01-WAR-2003	(TREMBLrel. 23, Created)			
Q8CF22	01-WAR-2003	(TREMBLrel. 23, Last sequence update)			
Q8CF22	01-OCT-2003	(TREMBLrel. 25, Last annotation update)			
Q8CF22	01-OCT-2003	Similar to subtilisin-like endoprotease (Fragment).			
Q8CF22	Mus musculus	(Mouse)			
Q8CF22	Eukaryota:	Metazoa;			
Q8CF22	Mammalia;	Eutheria;			
Q8CF22	Mammalia;	Rodentia;			
Q8CF22	NCBI_TaxId=10090;				
Q8CF22	[1]	SEQUENCE FROM N.A.			
Q8CF22	STRAIN=CZECH II;				
Q8CF22	STRAIN=CZECH RC				
Q8CF22	Strausberg R.;				
Q8CF22	Submitted (SEP-2002)	to the EMBL/GenBank/DBJ databases.			
Q8CF22	EMBL; BC037450;	AAH37450.1; --			
Q8CF22	GO; GO:0005489;	F:electron transporter activity; IEA.			
Q8CF22	GO; GO:0008233;	F:peptidase activity; IEA.			
Q8CF22	GO; GO:0004289;	F:subtilase activity; IEA.			
Q8CF22	GO; GO:0006118;	P:electron transport; IEA.			
Q8CF22	GO; GO:0006508;	P:proteolysis and peptidolysis; IEA.			
Q8CF22	InterPro; IPR00345;	CycC_heme_BS.			
Q8CF22	InterPro; IPR006212;	Purin repeat.			
Q8CF22	InterPro; IPR009030;	Grow_fac_recep.			
Q8CF22	InterPro; IPR006210;	IRGP.			
Q8CF22	InterPro; IPR000209;	Peptidase_S8			
Q8CF22	InterPro; IPR002884;	Peptidase_S8B.			
Q8CF22	InterPro; IPR007087;	Znf_C2H2.			
Q8CF22	Pfam; PF00082;	Peptidase_S8; 1.			
Q8CF22	Pfam; PF01483;	P_protein; 1.			
Q8CF22	PRINTS; PR00723;	SUBTILISIN.			
Q8CF22	ProDom; PD000717;	p_domain; 1.			
Q8CF22	SMART; SM00181;	EGF; 5.			
Q8CF22	SMART; SM00261;	FU; 5.			
Q8CF22	PROSITE; PS00190;	CYTOCHROME_C; 1.			
Q8CF22	PROSITE; PS00136;	SUBTILASE_ASP; 1.			
Q8CF22	PROSITE; PS00137;	SUBTILASE_HIS; 1.			
Q8CF22	PROSITE; PS00138;	SUBTILASE_SER; 1.			
Q8CF22	PROSITE; PS00028;	ZINC_FINGER_C2H2_1; 1.			
Q8CF22	Protease.				
Q8CF22	NON_TER				
Q8CF22	SEQUENCE	826 AA; 91553 MW; 61BCGB49F6F42AA0			
Q8CF22	CRC64;				

Query Match	11.18;	Score	171;	DB	11;	Length	826;
Best Local Similarity	22.97;	Pred. No.	5.8e-07;				
Matches	55;	Conservative	25;	Mismatches	88;	Indels	72;
						Gaps	12;

QY	18	YIGSQNASRRRRMRHPNVISQCGGCATCSQSYN--GCUSCKRPLRPFLVIERIGMKQIGV	75
			:
			:
DB	594	YFGDAARRCRR-----CHKGCETCTGRSPAQLCSR-RGPFY-----HHQBTNT	636
			:
OY	76	CLASSCPSGYGTRYTPINKTKCKVKVDCDTCPNK-NPCTCKSGFYHLGKCLDSCPGLE	134

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 16:54:09 ; Search time 49.1845 Seconds
(without alignments)
1562.545 Million cell updates/sec

Title: US-09-894-912A-34
Perfect score: 1510
Sequence: 1 MHRLISWLFILFMFYIG.....QQKKRVQDKSVSVSTVH 272

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1510	100.0	272	4 AAB99220	Aab99220 Human thr
2	1510	100.0	272	4 AAM78328	Aam78328 Human pro
3	1510	100.0	272	4 AAE13168	Aae13168 Human ste
4	1510	100.0	272	4 AAE13150	Aae13150 Human ste
5	1510	100.0	272	6 ABR62112	AbR62112 Human clo
6	1510	100.0	272	6 ABO44432	Abo44432 Human ste
7	1510	100.0	272	6 ABO44413	Abo44413 Human ste
8	1505	99.7	273	4 AAE13151	Aae13151 Human ste
9	1505	99.7	273	6 ABO44414	Abo44414 Human ste
10	1477	97.8	265	4 AAE13163	Aae13163 Human ste
11	1477	97.8	265	6 ABO44427	Abo44427 Human sec
12	1477	97.8	292	2 AAW85607	Aaw85607 Secreted
13	1477	97.8	292	4 AAE13170	Aae13170 Human SCR
14	1477	97.8	292	5 ABP61846	Abp61846 Human pol
15	1477	97.8	292	6 ABR62114	AbR62114 Human sec
16	1477	97.8	292	6 ABO44434	Abo44434 Human SCR
17	1457	96.5	292	4 AAB93875	Aab93875 Human pro
18	1400	92.7	251	4 AAE13153	Aae13153 Human mat
19	1400	92.7	251	6 ABO44417	Abo44417 Human ste
20	1315.5	87.1	279	4 AAE13167	Aae13167 Mouse ste
21	1315.5	87.1	279	6 ABO44431	Abo44431 Mouse ste
22	1183	78.3	239	6 ABUS2396	Abu52396 Human GPC
23	1111	73.6	195	6 ABUS2398	Abu52398 Human GPC
24	1107	73.3	195	6 ABUS2397	Abu52397 Human GPC
25	903	59.8	160	4 ABB11374	Abb11374 Human sec

26	903	59.8	160	4 AAM79312	Aam79312 Human pro
27	903	59.8	160	4 AAE13149	Aae13149 Human ste
28	903	59.8	160	6 ABO44415	Abo44415 Human ste
29	656	43.4	263	6 ABR62108	AbR62108 Secreted
30	656	43.4	263	6 ABR62115	AbR62115 Secreted
31	656	43.4	263	6 ABR58489	AbR58489 Human sec
32	656	43.4	263	7 ADE07919	AdE07919 Novel pro
33	648.5	42.9	243	6 ABR62110	AbR62110 Secreted
34	644	42.6	265	6 ABR62113	AbR62113 Mouse thr
35	638	42.3	229	4 AAE13162	Aae13162 Mouse thr
36	638	42.3	229	6 ABO44426	Abo44426 Mouse thr
37	589.5	39.0	243	6 AAE37115	Aae37115 Human sec
38	582.5	38.6	243	5 ABR62106	AbR62106 Secreted
39	582.5	38.6	243	6 ABR62107	AbR62107 Secreted
40	578.5	38.3	250	6 ABR62101	AbR62101 Secreted
41	578	38.3	222	6 ABR62102	AbR62102 Secreted
42	545	36.1	222	6 ABR62102	AbR62102 Secreted
43	540.5	35.8	229	6 ABR62102	AbR62102 Secreted
44	472	31.3	234	7 ADB76146	Adb76146 Novel hum
45	276	18.3	46	4 AAE13155	Aae13155 Human SCR

ALIGNMENTS

RESULT 1
AAB99220
ID AAB99220 standard; protein; 272 AA.
XX
AC AAB99220;
XX
DT 07-SEP-2001 (first entry)
XX
DE Human thrombospondin-30.
XX
KW Human; thrombospondin-30; cytostatic; anti-HIV; antiinflammatory;
KW malignant tumor; haemopathy; HIV infection; immunological disease;
KW inflammation disease.
XX
OS Homo sapiens.
XX
FN WO200140294-A1.
XX
PD 07-JUN-2001.
XX
PF 27-NOV-2000; 2000WO-CN000494.
XX
PR 29-NOV-1999; 99CN-00124148.
XX
PA (BIOR-) BIORAD GENE DEV LTD SHANGHAI.
XX
PI Mao Y, Xie Y;
XX
DR MPI; 2001-397948/42.
XX
DR N-PSDB; AAB45131.
XX
PT Human thrombospondin-30 and polynucleotide is useful in diagnosis and
PT treatment of, e.g., malignant tumor, hemopathy, HIV infection,
PT immunological diseases and various inflammatory diseases.
XX
PS Claim 1; Page 27-28; 33pp; Chinese.
XX
CC The present sequence is the protein sequence for human thrombospondin-30.
CC Thrombospondin-30 protein and coding sequence are useful in the diagnosis
CC and treatment of malignant tumor, haemopathy, HIV infection,
CC immunological diseases and various inflammation diseases. In addition
CC thrombospondin-30 protein may be used for screening mimics, agonists,
CC antagonists or inhibitors, or for use in peptide fingerprinting
CC identification. The thrombospondin-30 coding sequence may be used as
CC primers for nucleic acid amplification reaction or as probes for
CC hybridisation reaction, or in producing gene chips or microarrays
SQ Sequence 272 AA;

```

Query Match      100.0%; Score 1510; DB 4; Length 272;
Best Local Similarity 100.0%; Pred. No. 1e-110;
Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHLRLISLWFLIILNFMFMEYIGSONASRGRRORRHPNPNVSGCGGCATCSDYNGCLSCCKPR 60
DB 1 MHLRLISLWFLIILNFMFMEYIGSONASRGRRORRHPNPNVSGCGGCATCSDYNGCLSCCKPR 60

QY 61 LFPALERIGMKQIGVCLSSCPGSGYGYTRYDPDINKCTCKADCDTCFNNKFNCTCKSGFY 120
DB 61 LFPALERIGMKQIGVCLSSCPGSGYGYTRYDPDINKCTCKADCDTCFNNKFNCTCKSGFY 120

QY 121 HLKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPSPCTKKGTCGPKGTETRVREII 180
DB 121 HLKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPSPCTKKGTCGPKGTETRVREII 180

QY 181 QHPSAKGNLCPPTNETRKTCTVQRKKCKGGRGKGRERKRRKPNKGSKEAIPDSKSLES 240
DB 181 QHPSAKGNLCPPTNETRKTCTVQRKKCKGGRGKGRERKRRKPNKGSKEAIPDSKSLES 240

QY 241 SKETPEQRENKQOKKQKRVQDKQKSVSVTVH 272
DB 241 SKETPEQRENKQOKKQKRVQDKQKSVSVTVH 272

RESULT 2
AAW78328
ID AAW78328 standard; protein; 272 AA.
AC AAW78328;
XX
XX 06-NOV-2001 (first entry)
DE Human protein SEQ ID NO 990.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation.
XX
XX Homo sapiens.
XX
XX W0200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US004098.
XX
XX 03-FEB-2000; 2000US-00496914.
XX
XX 27-APR-2000; 2000US-00560875.
XX
XX 20-JUN-2000; 2000US-00598075.
XX
XX 19-JUL-2000; 2000US-00620325.
XX
XX 01-SEP-2000; 2000US-00654936.
XX
XX 15-SEP-2000; 2000US-00663561.
XX
XX 20-OCT-2000; 2000US-00693325.
XX
XX 30-NOV-2000; 2000US-00728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
XX Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren P, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
XX
XX N-PSDB; AAK51461.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
XX in diagnosis and gene therapy.
XX
XX Claim 20; Page 3214-3215; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the

```

```

CC encoded polypeptides (AAW78323-AAW80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAW80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
XX Sequence 272 AA;
XX
XX Query Match      100.0%; Score 1510; DB 4; Length 272;
XX Best Local Similarity 100.0%; Pred. No. 1e-110;
XX Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MHLRLISLWFLIILNFMFMEYIGSONASRGRRORRHPNPNVSGCGGCATCSDYNGCLSCCKPR 60
XX DB 1 MHLRLISLWFLIILNFMFMEYIGSONASRGRRORRHPNPNVSGCGGCATCSDYNGCLSCCKPR 60
XX
XX QY 61 LFPALERIGMKQIGVCLSSCPGSGYGYTRYDPDINKCTCKADCDTCFNNKFNCTCKSGFY 120
XX DB 61 LFPALERIGMKQIGVCLSSCPGSGYGYTRYDPDINKCTCKADCDTCFNNKFNCTCKSGFY 120
XX
XX QY 121 HLKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPSPCTKKGTCGPKGTETRVREII 180
XX DB 121 HLKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPSPCTKKGTCGPKGTETRVREII 180
XX
XX QY 181 QHPSAKGNLCPPTNETRKTCTVQRKKCKGGRGKGRERKRRKPNKGSKEAIPDSKSLES 240
XX DB 181 QHPSAKGNLCPPTNETRKTCTVQRKKCKGGRGKGRERKRRKPNKGSKEAIPDSKSLES 240
XX
XX QY 241 SKETPEQRENKQOKKQKRVQDKQKSVSVTVH 272
XX DB 241 SKETPEQRENKQOKKQKRVQDKQKSVSVTVH 272
XX
XX RESULT 3
XX AAE13168
XX ID AAE13168 standard; protein; 272 AA.
XX
XX AC AAE13168;
XX
XX XX 28-JAN-2002 (first entry)
XX
XX XX Human stem cell growth factor-like protein #4.
XX
XX Human; stem cell growth factor-like protein; antiinflammatory; neutropic;
XX neuroprotective; vulnary; cytostatic; anticonvulsant; immunostimulant;
XX vasotropic; virucide; dermatological; tranquillisier; cerebroprotective;
XX osteopathic; immunodeficiency syndrome; chronic granulomatous disease;
XX duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;
XX acquired immune deficiency syndrome; agammaglobulinaemia; thalassaemia;
XX Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;
XX adrenal white matter degeneration; anaemia; neurodegenerative disease;
XX Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;
XX severe combined immunodeficiency; immune disorder; autoimmune disorder;
XX multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
XX autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;
XX supporting factor for the proliferation of stem cell.
XX
XX Homo sapiens.
XX
XX OS
XX Key Location/Qualifiers
XX Peptide 1..21
XX Protein /label= Signal_peptide
XX /note= "Human mature stem cell growth factor-like
XX protein"

```


transplantation or cord blood transplantation for treating a variety of diseases such as immunodeficiency syndrome, chronic granulomatous disease, duplicated immunodeficiency syndrome, agammaglobulinaemia, Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS), thalassemia, haemolytic anaemia due to enzyme defect, congenital anaemia such as sickle cell anaemia, Gaucher's disease, lysosomal storage diseases such as mucopolysaccharidosis, adrenal white matter degeneration, a variety of cancer and tumours. Proteins of the invention are useful for treating diseases such as Parkinson's disease, Alzheimer's disease and other neurodegenerative diseases, thrombocytopaenia, immune deficiencies and disorders such as severe combined immunodeficiency (SCID) and autoimmune disorders such as multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary inflammation. Sequences of the invention are also useful in gene therapy. The present sequence is stem cell growth factor-like protein from human

XX
SQ

Query Match 100.0%; Score 1510; DB 4; Length 272;
Best Local Similarity 100.0%; Pred. No. 1e-110;
Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHLRLISWLFILFNFMEYIGSQNASRGRQRMRHPNVSQCGGCATCSDYNGCLSKPR 60
DB 1 MHLRLISWLFILFNFMEYIGSQNASRGRQRMRHPNVSQCGGCATCSDYNGCLSKPR 60
QY 61 LFPALERIGMKQIGVCLSSCPGYYTRYDPDINKTKCKADCTCFNKNFCTCKSGFYL 120
DB 61 LFPALERIGMKQIGVCLSSCPGYYTRYDPDINKTKCKADCTCFNKNFCTCKSGFYL 120
QY 121 HLGKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPSPCTKKGKTCGFRGTETRVREII 180
DB 121 HLGKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPSPCTKKGKTCGFRGTETRVREII 180
QY 181 QHPSAKGNLCPTNETRKTCTVQRKCKQGERGKGRERKRPKNKGSKEAIPDSKSL 240
DB 181 QHPSAKGNLCPTNETRKTCTVQRKCKQGERGKGRERKRPKNKGSKEAIPDSKSL 240
QY 241 SKEIPEORENKQOKRKVKQDKOKSVSVTVH 272
DB 241 SKEIPEORENKQOKRKVKQDKOKSVSVTVH 272

RESULT 5

ID ABR62112
AC ABR62112 standard; protein; 272 AA.

XX
AC ABR62112;

XX
DT 18-AUG-2003 (first entry)

XX
DE Human clone 1 thrombospondin protein #23.

XX
KW Human; secreted; stem cell growth factor; cytostatic; haemostatic; neuroprotective; immunostimulant; leukaemia; haemophilia; cancer; degenerative disease; Alzheimer's disease; food supplement; immunological disorder; thrombospondin.

XX
OS Homo sapiens.

XX
PN WO2003029405-A2.

XX
PD 10-APR-2003.

XX
PF 30-AUG-2002; 2002WO-US027746.

XX
PR 30-AUG-2001; 2001US-0316368P.

XX
PR 10-DEC-2001; 2001US-0339739P.

XX
PR 19-APR-2002; 2002US-00125852.

XX
PA (HYSB-) HYSBQ INC.

XX
Tang YT;

XX

DR WPI; 2003-381616/36.

XX
PT New stem cell growth factor-like polypeptides and polynucleotides, useful for treating e.g. leukemia, hemophilia and degenerative diseases like Alzheimer's disease, and for inducing immune response.

XX
PT Disclosure; Fig 1; 151pp; English.

XX
CC The invention relates to new stem cell growth factor-like polypeptides and polynucleotides. The stem cell growth factor-like polypeptides and polynucleotides are useful for inducing differentiation of embryonic and adult stem cells to give rise to different cell types, for treating e.g. leukaemia, haemophilia and degenerative diseases like Alzheimer's disease. They are also useful for generating new tissues and organs that may aid patients in need of transplanted tissues. The polynucleotides are useful as hybridisation probes, oligomers or primers for PCR, for chromosome and gene mapping, in recombinantly producing protein, in generating antisense DNA or RNA, in diagnostics as expressed sequence tags for identifying expressed genes, and for inducing immune response. The polypeptides are useful for generating antibodies that specifically bind the polypeptide, as molecular weight markers, and as a food supplement (e.g. protein or amino acid supplement, and as a carbon, nitrogen or carbohydrate source). Compositions comprising the polypeptides or polynucleotides are useful for the diagnosis, treatment or prevention of cancers, and other immunological disorders. The current sequence represents a human clone 1 thrombospondin protein

XX
SQ Sequence 272 AA;

Query Match 100.0%; Score 1510; DB 6; Length 272;
Best Local Similarity 100.0%; Pred. No. 1e-110;
Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHLRLISWLFILFNFMEYIGSQNASRGRQRMRHPNVSQCGGCATCSDYNGCLSKPR 60

DB 1 MHLRLISWLFILFNFMEYIGSQNASRGRQRMRHPNVSQCGGCATCSDYNGCLSKPR 60

QY 61 LFPALERIGMKQIGVCLSSCPGYYTRYDPDINKTKCKADCTCFNKNFCTCKSGFYL 120

DB 61 LFPALERIGMKQIGVCLSSCPGYYTRYDPDINKTKCKADCTCFNKNFCTCKSGFYL 120

QY 121 HLGKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPSPCTKKGKTCGFRGTETRVREII 180

DB 121 HLGKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPSPCTKKGKTCGFRGTETRVREII 180

QY 181 QHPSAKGNLCPTNETRKTCTVQRKCKQGERGKGRERKRPKNKGSKEAIPDSKSL 240

DB 181 QHPSAKGNLCPTNETRKTCTVQRKCKQGERGKGRERKRPKNKGSKEAIPDSKSL 240

QY 241 SKEIPEORENKQOKRKVKQDKOKSVSVTVH 272

DB 241 SKEIPEORENKQOKRKVKQDKOKSVSVTVH 272

RESULT 6

ID ABO44432
AC ABO44432 standard; protein; 272 AA.

XX
AC ABO44432;

XX
DT 30-SEP-2003 (first entry)

XX
DE Human stem cell growth factor-like protein, SCR 1 #4.

XX
KW Stem cell growth factor-like protein; antianemic; anti-HIV; SCR-1; immunostimulant; vulnery; haematopoietic stem cell; gene therapy; supporting factor for proliferation of stem cells; wound healing; haematopoietic progenitor cell; stromal cell; AIDS; thalassemia; bone marrow transplantation; cord blood transplantation; chronic granulomatous disease; duplicated immunodeficiency syndrome; agammaglobulinaemia; Wiskott-Aldrich syndrome; haemolytic anaemia; congenital anaemia; sicklaemia; Gaucher's disease; morphogenesis;

XX The patent discloses novel stem cell growth factor-like proteins and
 CC polynucleotides encoding them. Proteins of the invention are also known
 CC as supporting factor for the proliferation of stem cells (SCR-1). Stem
 CC cell growth factor-like proteins are useful for supporting proliferation
 CC or survival of a stem cell or germ cell which is preferably primordial
 CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem
 CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent
 CC cell. The haematopoietic progenitor cell cultured using stem cell growth
 CC factor-like proteins can replace as a graft for the bone marrow
 CC transplantation or cord blood transplantation for treating a variety of
 CC diseases such as immunodeficiency syndrome, chronic granulomatous
 CC disease, duplicated immunodeficiency syndrome, agammaglobulinaemia,
 CC Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),
 CC thalassaemia, haemolytic anaemia due to enzyme defect, congenital anaemia
 CC such as sickle cell anaemia, Gaucher's disease, lysosomal storage
 CC diseases such as mucopolysaccharidosis, adrenal white matter
 CC degeneration, a variety of cancer and tumours. Proteins of the invention
 CC are useful for treating diseases such as Parkinson's disease, Alzheimer's
 CC disease and other neurodegenerative diseases, thrombocytopaenia, immune
 CC deficiencies and disorders such as severe combined immunodeficiency
 CC (SCID) and autoimmune disorders such as multiple sclerosis, systemic
 CC lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary
 CC inflammation. Sequences of the invention are also useful in gene therapy.
 CC The present sequence is stem cell growth factor-like protein from human
 CC Sequence 273 AA;
 XX Query Match 99.7%; Score 1505; DB 4; Length 273;
 XX Best Local Similarity 100.0%; Pred. No. 2.5e-110;
 XX Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 HRLISLWLFILNMFYIGSNASRGRRMRMRHNPVSGGGGATGSDYNGCLSCPKRL 61
 DB 3 HRLISLWLFILNMFYIGSNASRGRRMRMRHNPVSGGGGATGSDYNGCLSCPKRL 62
 QY 62 FFALEIRIGMKOIGVCLSSCPGGYGTGTPDINKCTCKACDCTCFNKNFTCKCKSGYLH 121
 DB 63 FFALEIRIGMKOIGVCLSSCPGGYGTGTPDINKCTCKACDCTCFNKNFTCKCKSGYLH 122
 QY 122 LGKCLDNCPLGANNHMTVCVSIHCVSEFWNPSPCTKKGKTCGFKGTETVRVRIIQ 181
 DB 123 LGKCLDNCPLGANNHMTVCVSIHCVSEFWNPSPCTKKGKTCGFKGTETVRVRIIQ 182
 QY 182 HPSAKGNLCPPTNETRKTCTVQRKKCKGKRGKGRKRKKPKNGSKGAIPDSKLSLESS 241
 DB 183 HPSAKGNLCPPTNETRKTCTVQRKKCKGKRGKGRKRKKPKNGSKGAIPDSKLSLESS 242
 QY 242 KEIPEQRENKQKKRKVKQDKSVSVSTVH 272
 DB 243 KEIPEQRENKQKKRKVKQDKSVSVSTVH 273
 RESULT 9
 ABO44414
 ID ABO44414 standard; protein; 273 AA.
 XX AC ABO44414;
 XX 30-SEP-2003 (first entry)
 XX Human stem cell growth factor-like protein, SCR 1 #2.
 XX Stem cell growth factor-like protein; antianemic; anti-HIV; SCR-1;
 XX immunostimulant; vulnary; haematopoietic stem cell; gene therapy;
 XX supporting factor for proliferation of stem cells; wound healing;
 XX haematopoietic progenitor cell; stromal cell; AIDS; thalassaemia;
 XX bone marrow transplantation; cord blood transplantation;
 XX chronic granulomatous disease; duplicated immunodeficiency syndrome;
 XX agammaglobulinaemia; Wiskott-Aldrich syndrome; haemolytic anaemia;
 XX congenital anaemia; sickle cell; Gaucher's disease; morphogenesis;
 XX epithelial cell growth; ovarian follicle development; nerve cell growth;
 XX cartilage remodeling; bone growth; immunosuppression; human.

XX Homo sapiens.
 XX OS US2003044792-A1.
 XX PN 06-MAR-2003.
 XX PD 28-JUN-2001; 2001US-00894912.
 XX PF 28-JUN-2000; 2000US-0215733P.
 XX PR 05-FEB-2001; 2001US-0266614P.
 XX PR 05-APR-2001; 2001US-0282397P.
 XX XX (TANG/) TANG Y T.
 XX PA (LABA/) LABAT I.
 XX PA (DRMA/) DRMANAC R T.
 XX PA (MIZE/) MIZE N.
 XX PA (NISH/) NISHIKAWA M.
 XX PA (CHAO/) CHAO C.
 XX XX Tang YT, Labat I, Drmanac RT, Mize N, Nishikawa M, Chao C;
 XX WPI; 2003-625403/59.
 XX DR N-PSDB; ACH04325.
 XX XX Novel isolated polypeptide having stem cell growth factor activity,
 XX PT useful for promoting wound healing, and as a medicine to proliferate or
 XX PT support human hematopoietic stem cells or human hematopoietic progenitor
 XX PT cells.
 XX PS Claim 23; Page 72-73; 96pp; English.
 XX CC The invention relates to an isolated stem cell growth factor-like
 CC polypeptide (referred as supporting factor for proliferation of stem
 CC cells (SCR-1)) from mouse or human, or its mature protein portion, or
 CC fragment, analogue, variant or derivative, that retains stem cell growth
 CC factor activity. Also included are an isolated polynucleotide encoding
 CC SCR-1 (or its mature protein portion, or fragment, analogue, variant or
 CC derivative, that retains stem cell growth factor activity, or the
 CC complement of the polynucleotide), an (expression) vector comprising the
 CC SCR-1 polynucleotide, a host cell genetically engineered to contain the
 CC SCR-1 polynucleotide in operative association with a regulatory sequence
 CC that controls expression of the polynucleotide in the host cell,
 CC preparation of the SCR-1 polypeptide, a polypeptide which is an
 CC expression product of the SCR-1 polynucleotide (the polypeptide having an
 CC activity to support proliferation or survival of haematopoietic stem cell
 CC or haematopoietic progenitor cell, with a proviso that C-terminal aa
 CC sequence does not comprise the aa sequence appearing as ABO44433), an
 CC isolated SCR-1 polypeptide with stem cell growth factor activity and
 CC lacking any 10 consecutive aa from ABO44430, an isolated polypeptide
 CC with stem cell growth factor activity having at least an aa sequence
 CC appearing as ABO44428 and ABO44429, a culture medium comprising the SCR-1
 CC polypeptide to maintain survival of or promote proliferation of a stem
 CC cell or germ cell, an anti-SCR-1 antibody, a nucleic acid array
 CC comprising the SCR-1 polynucleotide or a unique segment of the SCR-1
 CC polynucleotide attached to a surface, a stromal cell genetically
 CC engineered to express the SCR-1 polypeptide to support proliferation or
 CC survival of a stem cell or germ cell and an implant comprising a cell
 CC genetically engineered to express the SCR-1 polypeptide. The SCR-1
 CC proliferation or survival of a stem cell or germ cell. The SCR-1
 CC polypeptide is useful for identifying a compound that binds to the SCR-1
 CC polypeptide and for maintaining survival of or promoting proliferation of
 CC a stem cell, a germ cell, a haematopoietic stem cell or a haematopoietic
 CC progenitor cell. The SCR-1 polypeptide is useful for promoting wound
 CC healing. The human haematopoietic stem cell or human haematopoietic
 CC progenitor cell culture using the SCR-1 polypeptide can replace as a
 CC graft for the conventional bone marrow transplantation or cord blood
 CC transplantation. The transplantation of haematopoietic stem cells can be
 CC employed as a therapy for treating diseases such as chronic granulomatous
 CC diseases, duplicated immunodeficiency syndrome, agammaglobulinaemia,
 CC Wiskott-Aldrich syndrome, AIDS, etc., thalassaemia, haemolytic anaemia
 CC due to enzyme defect, congenital anaemia such as sickle cell, Gaucher's
 CC disease etc. the SCR-1 polypeptide is useful for cell growth and

CC morphogenesis, including tissue specific stem cell growth, epithelial
 CC cell growth and regulation, ovarian follicle development, promoting nerve
 CC cell growth, sustaining neuronal populations, cartilage remodeling, bone
 CC growth and immunosuppression. The present sequence is a Human SCR-1
 CC protein
 XX
 SQ Sequence 273 AA;

Query Match 99.7%; Score 1505; DB 6; Length 273;
 Best Local Similarity 100.0%; Pred. No. 2.5e-110;
 Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HRLISLWLFILNFMVEYIGSNASGRQRMRHNPVSGCGCATCDYNGCLSKPRL 61
 Db 3 HRLISLWLFILNFMVEYIGSNASGRQRMRHNPVSGCGCATCDYNGCLSKPRL 62
 Qy 62 FFALERIGMKQIGVCLSSCPGYYGTRYPDINKCTCKKADCDTCFNNFCTCKSGPYLH 121
 Db 63 FFALERIGMKQIGVCLSSCPGYYGTRYPDINKCTCKKADCDTCFNNFCTCKSGPYLH 122
 Qy 122 LGKCLDNCPEGLEANNHTMBCVSIHVCEVSEWNPSPCTKGGTCGPKRGTTETVRREII 181
 Db 123 LGKCLDNCPEGLEANNHTMBCVSIHVCEVSEWNPSPCTKGGTCGPKRGTTETVRREII 182
 Qy 182 HPSAKGNLCPTNETRKTCTVQRKKCKGRGKGRKRKPKNGESKEALPDSKSLESS 241
 Db 183 HPSAKGNLCPTNETRKTCTVQRKKCKGRGKGRKRKPKNGESKEALPDSKSLESS 242
 Qy 242 KEIPEQRENKQKKRKVKQDKQKSVSVTVH 272
 Db 243 KEIPEQRENKQKKRKVKQDKQKSVSVTVH 273

RESULT 10
 AAEL13163
 ID AAEL13163 standard; protein; 265 AA.

AC AAEL13163;
 XX
 XX 28-JAN-2002 (first entry)
 XX Human secreted protein from clone DA228_6.

XX Human; stem cell growth factor-like protein; antiinflammatory; neutropic;
 KW neuroprotective; vulnerable; cytostatic; anticonvulsant; immunostimulant;
 KW vasotropic; virucide; dermatological; tranquilliser; cerebroprotective;
 KW osteopathic; immunodeficiency syndrome; chronic granulomatous disease;
 KW duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;
 KW acquired immune deficiency syndrome; agammaglobulinaemia; thalassaemia;
 KW Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;
 KW adrenal white matter degeneration; anaemia; neurodegenerative disease;
 KW Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;
 KW severe combined immunodeficiency; immune disorder; autoimmune disease;
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
 KW autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;
 KW supporting factor for the proliferation of stem cell; secreted protein.

XX Homo sapiens.
 XX W0200177169-A2.
 XX 18-OCT-2001.

XX 05-APR-2001; 2001WO-US011208.
 XX 05-APR-2000; 2000US-00543774.
 PR 28-JUN-2000; 2000US-0215733P.
 PR 09-JAN-2001; 2001US-00757562.
 PR 05-FEB-2001; 2001US-0266614P.

XX (HYSB-) HYSEQ INC.
 PA (KIRI) KIRIN BEER KK.

XX

PI Tang TY, Labat I, Tillinghast JS, Sinku A, Liu C, Drmanac RT;
 PI Stache-Crain B, Dickson M, Mize NK, Nishikawa M;
 XX WPI; 2001-657166/75.

XX Novel stem cell growth factor like polypeptides and polynucleotides for
 PT identifying modulators useful for treating diseases such as Alzheimer's
 PT disease, cancer, rheumatoid arthritis, osteoporosis.

XX Disclosure; Fig 3; 232pp; English.

XX The patent discloses novel stem cell growth factor-like proteins and
 CC polynucleotides encoding them. Proteins of the invention are also known
 CC as supporting factor for the proliferation of stem cells (SCR-1). Stem
 CC cell growth factor-like proteins are useful for supporting proliferation
 CC or survival of a stem cell or germ cell which is preferably primordial
 CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem
 CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent
 CC cell. The haematopoietic progenitor cell cultured using stem cell growth
 CC factor-like proteins can replace as a graft for the bone marrow
 CC transplantation or cord blood transplantation for treating a variety of
 CC diseases such as immunodeficiency syndrome, chronic granulomatous
 CC disease, duplicated immunodeficiency syndrome, agammaglobulinaemia,
 CC Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),
 CC thalassaemia, haemolytic anaemia due to enzyme defect, congenital anaemia
 CC such as sickle cell anaemia, Gaucher's disease, lysosomal storage
 CC diseases such as mucopolysaccharidosis, adrenal white matter
 CC degeneration, a variety of cancer and tumours. Proteins of the invention
 CC are useful for treating diseases such as Parkinson's disease, Alzheimer's
 CC disease and other neurodegenerative diseases, thrombocytopaenia, immune
 CC deficiencies and disorders such as severe combined immunodeficiency
 CC (SCID) and autoimmune disorders such as multiple sclerosis, systemic
 CC lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary
 CC inflammation. Sequences of the invention are also useful in gene therapy.
 CC The present sequence is human secreted protein from clone DA228_6
 XX
 SQ Sequence 265 AA;

Query Match 97.8%; Score 1477; DB 4; Length 265;
 Best Local Similarity 100.0%; Pred. No. 3.8e-108;
 Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MELRLISLWLFILNFMVEYIGSNASGRQRMRHNPVSGCGCATCDYNGCLSKPRL 60
 Db 1 MELRLISLWLFILNFMVEYIGSNASGRQRMRHNPVSGCGCATCDYNGCLSKPRL 60
 Qy 61 LFFALERIGMKQIGVCLSSCPGYYGTRYPDINKCTCKKADCDTCFNNFCTCKSGPYL 120
 Db 61 LFFALERIGMKQIGVCLSSCPGYYGTRYPDINKCTCKKADCDTCFNNFCTCKSGPYL 120
 Qy 121 HLKCLDNCPEGLEANNHTMBCVSIHVCEVSEWNPSPCTKGGTCGPKRGTTETVRREII 180
 Db 121 HLKCLDNCPEGLEANNHTMBCVSIHVCEVSEWNPSPCTKGGTCGPKRGTTETVRREII 180
 Qy 181 QHPSAKGNLCPTNETRKTCTVQRKKCKGRGKGRKRKPKNGESKEALPDSKSLES 240
 Db 181 QHPSAKGNLCPTNETRKTCTVQRKKCKGRGKGRKRKPKNGESKEALPDSKSLES 240
 Qy 241 SKEIPEQRENKQKKRKVKQDKQKS 265
 Db 241 SKEIPEQRENKQKKRKVKQDKQKS 265

RESULT 11
 ABO44427
 ID ABO44427 standard; protein; 265 AA.

XX ABO44427;

XX 30-SEP-2003 (first entry)

XX Human secreted protein clone DA228_6.

XX

KW Stem cell growth factor-like protein; antianemic; anti-HIV; SCR-1;
 KW immunostimulant; vulnary; haematopoietic stem cell; gene therapy;
 KW supporting factor for proliferation of stem cells; wound healing;
 KW haematopoietic progenitor cell; stromal cell; AIDS; thalassemia;
 KW bone marrow transplantation; cord blood transplantation;
 KW chronic granulomatous disease; duplicated immunodeficiency syndrome;
 KW agammaglobulinaemia; Wiskott-Aldrich syndrome; haemolytic anaemia;
 KW congenital anaemia; sickle cell disease; Gaucher's disease; morphogenesis;
 KW epithelial cell growth; ovarian follicle development; nerve cell growth;
 KW cartilage remodeling; bone growth; immunosuppression; human.

XX Homo sapiens.

XX US2003044792-A1.

XX 06-MAR-2003.

XX 28-JUN-2001; 2001US-00894912.

XX 28-JUN-2000; 2000US-0215733P.

XX 03-FEB-2001; 2001US-0286614P.

XX 03-APR-2001; 2001US-0282397P.

XX (TANG/) TANG Y T.

XX (LABA/) LABAT I.

XX (DRMA/) DRMANAC R T.

XX (MIZE/) MIZE N.

XX (NISH/) NISHIKAWA M.

XX (CHAO/) CHAO C.

XX Tang YT, Labat I, Drmanac RT, Mize N, Nishikawa M, Chao C;

XX WPI; 2003-625403/59.

XX Novel isolated polypeptide having stem cell growth factor activity,

XX useful for promoting wound healing, and as a medicine to proliferate or

XX support human hematopoietic stem cells or human hematopoietic progenitor

XX cells.

XX Disclosure; Fig 3; 96pp; English.

XX The invention relates to an isolated stem cell growth factor-like
 CC polypeptide (referred as supporting factor for proliferation of stem
 CC cells (SCR-1)) from mouse or human, or its mature protein portion, or
 CC fragment, analogue, variant or derivative, that retains stem cell growth
 CC factor activity. Also included are an isolated polynucleotide encoding
 CC SCR-1 (or its mature protein portion, or fragment, analogue, variant or
 CC derivative, that retains stem cell growth factor activity, or the
 CC complement of the polynucleotide), an (expression) vector comprising the
 CC SCR-1 polynucleotide, a host cell genetically engineered to contain the
 CC SCR-1 polynucleotide in operative association with a regulatory sequence
 CC that controls expression of the polynucleotide in the host cell,
 CC preparation of the SCR-1 polypeptide, a polypeptide which is an
 CC expression product of the SCR-1 polynucleotide (the polypeptide having an
 CC activity to support proliferation or survival of haematopoietic stem cell
 CC or haematopoietic progenitor cell, with a proviso that C-terminal aa
 CC sequence does not comprise the aa sequence appearing as ABO44433), an
 CC isolated SCR-1 polypeptide with stem cell growth factor activity and
 CC lacking any 10 consecutive aas from ABO44430, an isolated polypeptide
 CC with stem cell growth factor activity having at least an aa sequence
 CC appearing as ABO44428 and ABO44429, a culture medium comprising the SCR-1
 CC polypeptide to maintain survival of or promote proliferation of a stem
 CC cell or germ cell, an anti-SCR-1 antibody, a nucleic acid array
 CC comprising the SCR-1 polynucleotide or a unique segment of the SCR-1
 CC polynucleotide attached to a surface, a stromal cell genetically
 CC engineered to express the SCR-1 polypeptide to support proliferation or
 CC survival of a stem cell or germ cell and an implant comprising a cell
 CC genetically engineered to express the SCR-1 polypeptide to support
 CC proliferation or survival of a stem cell or germ cell. The SCR-1
 CC polypeptide is useful for identifying a compound that binds to the SCR-1
 CC polypeptide and for maintaining survival of or promoting proliferation of
 CC a stem cell, a germ cell, a haematopoietic stem cell or a haematopoietic
 CC progenitor cell. The SCR-1 polypeptide is useful for promoting wound

CC healing. The human haematopoietic stem cell or human haematopoietic
 CC progenitor cell culture using the SCR-1 polypeptide can replace as a
 CC graft for the conventional bone marrow transplantation or cord blood
 CC transplantation. The transplantation of haematopoietic stem cells can be
 CC employed as a therapy for treating diseases such as chronic granulomatous
 CC diseases, duplicated immunodeficiency syndrome, agammaglobulinaemia,
 CC Wiskott-Aldrich syndrome, AIDS, etc., thalassemia, haemolytic anaemia
 CC due to enzyme defect, congenital anaemia such as sickle cell disease, Gaucher's
 CC disease etc. The SCR-1 polypeptide is useful for cell growth, and
 CC morphogenesis, including tissue specific stem cell growth, epithelial
 CC cell growth, and regulation, ovarian follicle development, promoting nerve
 CC cell growth, sustaining neuronal populations, cartilage remodeling, bone
 CC growth and immunosuppression. The present sequence is a protein
 CC homologous to Human SCR-1 protein

XX SQ Sequence 265 AA;

Query Match 97.8%; Score 1477; DB 6; Length 265;

Best Local Similarity 100.0%; Pred. No. 3.8e-108;

Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHLRLISLWLFILNFMFYIGSQNASRGRQRMRMPNVSGQGGCATSDYNGCLSKR 60

Db 1 MHLRLISLWLFILNFMFYIGSQNASRGRQRMRMPNVSGQGGCATSDYNGCLSKR 60

Qy 61 LPPALERIGMKOIGVCLSSCPSSGYGTRYPDINKCTKADCDTCFNKPFCTKSGFYL 120

Db 61 LPPALERIGMKOIGVCLSSCPSSGYGTRYPDINKCTKADCDTCFNKPFCTKSGFYL 120

Qy 121 HLGLCLDNCPGLEANNHMECVSIHVCEVSEWNPSPCTKGGKTCGPRGTETRVREII 180

Db 121 HLGLCLDNCPGLEANNHMECVSIHVCEVSEWNPSPCTKGGKTCGPRGTETRVREII 180

Qy 181 QHPSAKGNLCPTNTRKCTVQRKKCGKRGKGRKRKPKPNKGSKEALPDSKSLSS 240

Db 181 QHPSAKGNLCPTNTRKCTVQRKKCGKRGKGRKRKPKPNKGSKEALPDSKSLSS 240

Qy 241 SKBIPBQRENKQKKRKVKQDKQS 265

Db 241 SKBIPBQRENKQKKRKVKQDKQS 265

RESULT 12

AAW85607

ID AAW85607 standard; protein; 292 AA.

AC AAW85607;

DT 02-MAR-1999 (first entry)

DE Secreted protein clone da228_6.

KW Clone; secreted protein; protein factor; cytokine; lymphokine;

KW interferon; colony stimulating factor; CSF; interleukin; cloning;

KW tumour invasion; tumour suppression; immune boosting.

XX Homo sapiens.

XX WO9849302-A1.

XX 05-NOV-1998.

XX 24-APR-1998; 98WO-US008336.

XX 25-APR-1997; 97US-00845296.

XX 23-APR-1998; 98US-00065125.

XX (GENY) GENETICS INST INC.

XX Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;

XX Spaulding V, Agostino MJ;

XX WPI; 1999-024059/02.

DR N-PSDB; AAV831133.

XX New polynucleotides encoding secreted human proteins - are derived from

PT human foetal brain, adult brain, adult blood or placenta cDNA libraries,

PT useful, e.g. as potential immunomodulators.

XX Claim 8; Page 63-64; 104pp; English.

XX The nucleotide sequence (NS) of the full-length protein-coding sequence

CC of clones c1254 (AAV83132), da2286 (AAV83133), du4105 (AAV83134), eh801

CC (AAV83135), er3691 (AAV83136), fh1235 (AAV83137), fm601 (AAV83138) or

CC fr4732 (AAV83139). (all clones are deposited as ATCC 98415) and the

CC proteins they encode are predicted to have biological activities which

CC would make them suitable for treating, preventing or ameliorating medical

CC conditions in humans and animals for example, tumour suppression/invasion

CC activity, immune system boosting activity. The polynucleotides are also

CC believed to be useful for gene therapy

XX

SQ Sequence 292 AA;

Query Match 97.8%; Score 1477; DB 2; Length 292;

Best Local Similarity 100.0%; Pred. NO. 4.3e-108;

Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHRLISWLPILINFMVEYIGSNASRGRRORRHHPNVSGCGGCATCSDYNGCLSCKPR 60

DB 1 MHRLISWLPILINFMVEYIGSNASRGRRORRHHPNVSGCGGCATCSDYNGCLSCKPR 60

QY 61 LFPALERIGMKQIGVCLSSCPSGYGTRYDPDINKTKCKADCDTCFKNKFCCKSGFYL 120

DB 61 LFPALERIGMKQIGVCLSSCPSGYGTRYDPDINKTKCKADCDTCFKNKFCCKSGFYL 120

QY 121 HLKGLDNCPEGLEANNHTMECVSIHVCHVSEWNPSPCTKKGTCGFKGTETRVREII 180

DB 121 HLKGLDNCPEGLEANNHTMECVSIHVCHVSEWNPSPCTKKGTCGFKGTETRVREII 180

QY 181 QHPSAGNLCPPTNTRKCTVQRKKCKQGERKKGRKKRKNKPKGSKKAI PDSKSL 240

DB 181 QHPSAGNLCPPTNTRKCTVQRKKCKQGERKKGRKKRKNKPKGSKKAI PDSKSL 240

QY 241 SKEIPEQRENKQKKRKVDKQKS 265

DB 241 SKEIPEQRENKQKKRKVDKQKS 265

RESULT 13

ID AAE13170

XX AAE13170 standard; protein; 292 AA.

AC AAE13170;

XX

DT 28-JAN-2002 (first entry)

XX

DE Human SCR-1 related protein.

XX

XX Human; stem cell growth factor-like protein; antiinflammatory; nootropic;

XX neuroprotective; vulnary; cytostatic; anticonvulsant; immunostimulant;

XX vasotropic; virucide; dermatological; tranquiliser; cerebroprotective;

XX osteopathic; immunodeficiency syndrome; chronic granulomatous disease;

XX duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;

XX acquired immune deficiency syndrome; agammaglobulinaemia; thalasasaemia;

XX Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;

XX adrenal white matter degeneration; anaemia; neurodegenerative disease;

XX Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;

XX severe combined immunodeficiency; immune disorder; autoimmune disorder;

XX multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;

XX autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;

XX supporting factor for the proliferation of stem cell.

XX Unidentified.

OS

XX WO200177169-A2.

PN

XX

XX

PD 18-OCT-2001.

XX 05-APR-2001; 2001WO-US011208.

XX

XX 05-APR-2000; 2000US-00543774.

PR 28-JUN-2000; 2000US-0215733P.

PR 09-JAN-2001; 2001US-00757562.

PR 05-FEB-2001; 2001US-0266614P.

XX (HYSE-) HYSEQ INC.

PA (KIRI) KIRIN BEER KK.

XX Tang TY, Labat I, Tillinghast JS, Sinku A, Liu C, Drmanac RT;

XX Scache-Crain B, Dickson M, Mize NK, Nishikawa M;

PI WPI; 2001-657166/75.

DR N-PSDB; AAD21740.

XX

XX Novel stem cell growth factor like polypeptides and polynucleotides for

PT identifying modulators useful for treating diseases such as Alzheimer's

PT disease, cancer, rheumatoid arthritis, osteoporosis.

XX

PS Claim 27; Page 231-232; 232pp; English.

XX

CC The patent discloses novel stem cell growth factor-like proteins and

CC polynucleotides encoding them. Proteins of the invention are also known

CC as supporting factor for the proliferation of stem cells (SCR-1). Stem

CC cell growth factor-like proteins are useful for supporting proliferation

CC or survival of a stem cell or germ cell which is preferably primordial

CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem

CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent

CC cell. The haematopoietic progenitor cell cultured using stem cell growth

CC factor-like proteins can replace as a graft for the bone marrow

CC transplantation or cord blood transplantation for treating a variety of

CC diseases such as immunodeficiency syndrome, chronic granulomatous

CC disease, duplicated immunodeficiency syndrome, agammaglobulinaemia,

CC Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),

CC thalasasaemia, haemolytic anaemia due to enzyme defect, congenital anaemia

CC such as sickle cell anaemia, Gaucher's disease, lysosomal storage

CC diseases such as mucopolysaccharidosis, adrenal white matter

CC degeneration, a variety of cancer and tumours. Proteins of the invention

CC are useful for treating diseases such as Parkinson's disease, Alzheimer's

CC disease and other neurodegenerative diseases, thrombocytopaenia, immune

CC deficiencies and disorders such as severe combined immunodeficiency

CC (SCID) and autoimmune disorders such as multiple sclerosis, systemic

CC lupus erythematosus, rheumatoid arthritis, and autoimmune pulmonary

CC inflammation. Sequences of the invention are also useful in gene therapy.

CC The present sequence is human SCR-1 related protein

XX

SQ Sequence 292 AA;

Query Match 97.8%; Score 1477; DB 4; Length 292;

Best Local Similarity 100.0%; Pred. NO. 4.3e-108;

Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHRLISWLPILINFMVEYIGSNASRGRRORRHHPNVSGCGGCATCSDYNGCLSCKPR 60

DB 1 MHRLISWLPILINFMVEYIGSNASRGRRORRHHPNVSGCGGCATCSDYNGCLSCKPR 60

QY 61 LFPALERIGMKQIGVCLSSCPSGYGTRYDPDINKTKCKADCDTCFKNKFCCKSGFYL 120

DB 61 LFPALERIGMKQIGVCLSSCPSGYGTRYDPDINKTKCKADCDTCFKNKFCCKSGFYL 120

QY 121 HLKGLDNCPEGLEANNHTMECVSIHVCHVSEWNPSPCTKKGTCGFKGTETRVREII 180

DB 121 HLKGLDNCPEGLEANNHTMECVSIHVCHVSEWNPSPCTKKGTCGFKGTETRVREII 180

QY 181 QHPSAGNLCPPTNTRKCTVQRKKCKQGERKKGRKKRKNKPKGSKKAI PDSKSL 240

DB 181 QHPSAGNLCPPTNTRKCTVQRKKCKQGERKKGRKKRKNKPKGSKKAI PDSKSL 240

QY 241 SKEIPEQRENKQKKRKVDKQKS 265

Db 241 SKEIPEQENKQQKKRKVDKQKS 265

RESULT 14
ABP61846
ID ABP61846 standard; protein; 292 AA.
AC ABP61846;
XX
DT 04-OCT-2002 (first entry)
XX
DB Human polypeptide SEQ ID NO 200.
XX
KW Human; cytostatic; antirheumatic; antiarthritic; vulnary; analgesic;
KW antiinflammatory; antibacterial; immunosuppressive; antiparkinsonian;
KW neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;
KW anticancer; fungicide; antidiabetic; antiaschmatic; antiallergic;
KW immunostimulant; antiparasitic; secreted protein; transmembrane protein;
KW cytokine; cell proliferation; cell differentiation; autoimmune disease;
KW stem cell; growth factor; nervous system disease; neuropathy;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW osteoporosis; severe combined immunodeficiency; SCID; infection;
KW multiple sclerosis; rheumatoid arthritis; gene therapy.
XX
OS Homo sapiens.
XX
FN US2002065394-A1.
XX
PD 30-MAY-2002.
XX
PF 22-DEC-2000; 2000US-00745763.
XX
PR 18-MAR-1998; 98US-00040963.
XX
PA (JACO/) JACOBS K.
PA (MCCO/) MCCOY J.M.
PA (LAVA/) LAVALLIE E.R.
PA (COLL/) COLLINS-RACIE L.A.
PA (EVAN/) EVANS C.
PA (MERB/) MERBERG D.
PA (TREA/) TREACY M.
PA (SPAU/) SPAULDING V.
XX
PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Spaulding V;
XX
DR WPI: 2002-582343/62.
DR N-PSDB; ABQ92060.
XX
PT Novel secreted or transmembrane protein and polynucleotide encoding the
PT protein, useful for diagnosis and treatment of neurological disorders,
PT cancer, autoimmune diseases, bone disorders and lung or liver fibrosis.
XX
PS Claim 207; Page 203-204; 284pp; English.
XX
CC The invention relates to human secreted or transmembrane protein (I),
CC their fragments and is encoded by specific complementary deoxyribonucleic
CC acid (cDNA) inserts (II), where the protein is substantially free from
CC other mammalian proteins. (I) are useful for preventing, treating or
CC ameliorating a medical condition, especially immunological treatment or
CC prevention of tumours. (I) exhibits activity relating to angiogenesis,
CC cytokine, cell proliferation, cell differentiation, antiinflammatory,
CC stem cell growth factor activity and activin or inhibin-related
CC activities. (I) can be used to manipulate stem cells in culture to give
CC rise to neuroepithelial cells that can be used to augment or replace
CC cells damaged by illness, autoimmune disease, accidental damage or
CC genetic disorders. (I) induces the proliferation of neural cells and
CC regeneration of nerve and brain tissue and is useful for the treatment of
CC central and peripheral nervous system diseases and neuropathies, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis. (I) is involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis and is useful for treating myeloid
CC or lymphoid cell disorders, platelet disorders such as thrombocytopaenia

CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve
CC tissue growth and in tissue repair, healing of burns, incisions, ulcers,
CC for treating osteoporosis, osteoarthritis, bone degenerative disorders or
CC periodontal disease. (I) is also useful for gut protection or
CC regeneration and treatment of lung or liver fibrosis, reperfusion injury
CC in various tissues, various immune deficiencies and disorders including
CC severe combined immunodeficiency (SCID), bacterial or fungal infections,
CC autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis,
CC diabetes mellitus, myasthenia gravis, allergic reactions and conditions,
CC such as asthma or other respiratory problems. (II) is useful to express
CC recombinant protein, as markers for tissues in which the corresponding
CC protein is preferentially expressed and in gene therapy. The present
CC sequence is that of a polypeptide of the invention
XX

SEQ Sequence 292 AA;

Query Match 97.8%; Score 1477; DB 5; Length 292;
Best Local Similarity 100.0%; Pred. No. 4.3e-108;
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHLRLISWLFILNPMYIGSONASGRGRORRMHPNVSGCGGCATCSDYNGCLSKPR 60
DB 1 MHLRLISWLFILNPMYIGSONASGRGRORRMHPNVSGCGGCATCSDYNGCLSKPR 60
QY 61 LPEALERIGMKQIGVCLSSCPGSGYGYTRYPDINKCTKCKADCDTCFNKNFCTCKSGFYL 120
DB 61 LPEALERIGMKQIGVCLSSCPGSGYGYTRYPDINKCTKCKADCDTCFNKNFCTCKSGFYL 120
QY 121 HLGKCLDNCPEGLBANNHMECVSIHVCEVSEWNPWSPCTKKGKTCGPKRGTTETRVREII 180
DB 121 HLGKCLDNCPEGLBANNHMECVSIHVCEVSEWNPWSPCTKKGKTCGPKRGTTETRVREII 180
QY 181 QHPSAKGNLCPPTNETRKTCTVQRKKCKQGERGKGRERKRPKNKGSKEAIPDSKSLSS 240
DB 181 QHPSAKGNLCPPTNETRKTCTVQRKKCKQGERGKGRERKRPKNKGSKEAIPDSKSLSS 240
QY 241 SKEIPEQENKQQKKRKVDKQKS 265
DB 241 SKEIPEQENKQQKKRKVDKQKS 265

RESULT 15

ABR62114
ID ABR62114 standard; protein; 292 AA.
XX
AC ABR62114;
XX
DT 18-AUG-2003 (first entry)
XX
DE Human secreted protein clone da_288_6 #25.

XX Human; secreted; stem cell growth factor; cytostatic; haemostatic;
KW neuroprotective; immunostimulant; leukaemia; haemophilia; cancer;
KW degenerative disease; Alzheimer's disease; food supplement;
KW immunological disorder.

XX Homo sapiens.

XX WO2003029405-A2.

XX 10-APR-2003.

XX 30-AUG-2002; 2002WO-US027746.

XX 30-AUG-2001; 2001US-0316368P.

PR 10-DEC-2001; 2001US-0339739P.

PR 19-APR-2002; 2002US-00125852.

XX (HYSB-) HYSSEQ INC.

XX Tang YT;

DR WPI; 2003-381616/36.

XX New stem cell growth factor-like polypeptides and polynucleotides, useful
PT for treating e.g. leukemia, hemophilia and degenerative diseases like
PT Alzheimer's disease, and for inducing immune response.
XX
PS Disclosure; Fig 2; 151pp; English.
PS
XX The invention relates to new stem cell growth factor-like polypeptides
CC and polynucleotides. The stem cell growth factor-like polypeptides and
CC polynucleotides are useful for inducing differentiation of embryonic and
CC adult stem cells to give rise to different cell types, for treating e.g.
CC leukaemia, haemophilia and degenerative diseases like Alzheimer's
CC disease. They are also useful for generating new tissues and organs that
CC may aid patients in need of transplanted tissues. The polynucleotides are
CC useful as hybridisation probes, oligomers or primers for PCR, for
CC chromosome and gene mapping, in recombinantly producing protein, in
CC generating antisense DNA or RNA, in diagnostics as expressed sequence
CC tags for identifying expressed genes, and for inducing immune response.
CC The polypeptides are useful for generating antibodies that specifically
CC bind the polypeptide, as molecular weight markers, and as a food
CC supplement (e.g. protein or amino acid supplement, and as a carbon,
CC nitrogen or carbohydrate source). Compositions comprising the
CC polypeptides or polynucleotides are useful for the diagnosis, treatment
CC or prevention of cancers, and other immunological disorders. The current
CC sequence represents a human secreted protein clone da_288_6
XX
SQ Sequence 292 AA;
Query Match 97.8%; Score 1477; DB 6; Length 292;
Best Local Similarity 100.0%; Pred. No. 4.3e-108;
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHLRLISWLFIIILFMEYIGSQNASRGRORRHPNVSQCGGCATCSDYNGCLSCKPR 60
DB 1 MHLRLISWLFIIILFMEYIGSQNASRGRORRHPNVSQCGGCATCSDYNGCLSCKPR 60
QY 61 LFFALERIGMKOIGVCLSSCPGSGYGYTRYPDINKTKCKADCDTCFNKPFCTKCKSGPYL 120
DB 61 LFFALERIGMKOIGVCLSSCPGSGYGYTRYPDINKTKCKADCDTCFNKPFCTKCKSGPYL 120
QY 121 HLGKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPMSPTCKGKTCGFKRGTTETRVREII 180
DB 121 HLGKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPMSPTCKGKTCGFKRGTTETRVREII 180
QY 181 QHPSAKGNLCPTNETRKTCTVQRKKQKGERGKGRERKRKPKNKGESKEAIPDSKSLES 240
DB 181 QHPSAKGNLCPTNETRKTCTVQRKKQKGERGKGRERKRKPKNKGESKEAIPDSKSLES 240
QY 241 SKEIPEORENKQOKKRVQDKQS 265
DB 241 SKEIPEORENKQOKKRVQDKQS 265

Search completed: June 29, 2004, 16:59:53
Job time : 50.5178 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 16:57:05 ; Search time 14.1942 Seconds
(without alignment)
989.298 Million cell updates/sec

Title: US-09-894-912A-34

Perfect score: 1510

Sequence: 1 MHLRLISWLPILNPFMYIG.....QQKKRKVDKQKSVSVTVH 272

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A-COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B-COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A-COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B-COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS-COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	173.5	11.5	969	2	US-08-284-941-2
2	173.5	11.5	969	2	US-08-447-642-2
3	173.5	11.5	969	3	US-09-236-503-2
4	173.5	11.5	969	5	PCT-US93-02147A-2
5	162	10.7	799	2	US-08-525-940-23
6	162	10.7	799	2	US-08-976-838-23
7	162	10.7	881	2	US-08-525-940-21
8	162	10.7	881	2	US-08-976-838-21
9	162	10.7	915	2	US-08-525-940-18
10	162	10.7	915	2	US-08-976-838-18
11	162	10.7	915	4	US-09-214-555B-2
12	162	10.7	915	4	US-09-214-555B-7
13	161	10.7	288	1	US-08-368-852-15
14	158.5	10.5	379	4	US-09-907-794A-4
15	158.5	10.5	379	4	US-09-905-125A-4
16	158.5	10.5	379	4	US-09-902-775A-4
17	157.5	10.4	288	2	US-08-525-940-15
18	157.5	10.4	288	2	US-08-976-838-15
19	150.5	10.0	380	4	US-09-205-258-441
20	149	9.9	568	1	US-07-862-021B-14
21	149	9.9	568	5	PCT-US93-03164-14
22	148.5	9.8	2523	1	US-08-185-432-18
23	148.5	9.8	2523	4	US-08-899-232-3
24	147.5	9.8	802	1	US-07-862-021B-12
25	147.5	9.8	802	1	US-08-313-288B-12
26	147.5	9.8	802	5	PCT-US93-03164-12
27	146.5	9.7	807	4	US-09-132-769-1

28 146.5 9.7 807 4 US-09-640-173-186 Sequence 186, App
29 146.5 9.7 807 4 US-08-713-550-186 Sequence 186, App
30 146 9.7 1068 1 US-08-537-210A-2 Sequence 2, Appli
31 146 9.7 1068 3 US-09-113-825-2 Sequence 2, Appli
32 146 9.7 2556 1 US-08-185-432-17 Sequence 17, Appl
33 146 9.7 2556 4 US-08-899-232-2 Sequence 2, Appli
34 144 9.5 807 4 US-09-132-769-3 Sequence 3, Appli
35 143.5 9.5 807 1 US-07-862-021B-10 Sequence 10, Appl
36 143.5 9.5 807 1 US-08-313-288B-10 Sequence 10, Appl
37 143.5 9.5 807 4 US-09-132-769-5 Sequence 5, Appl
38 143.5 9.5 807 5 PCT-US93-03164-10 Sequence 103, App
39 139 9.2 366 3 US-08-857-076-103 Sequence 5, Appli
40 139 9.2 486 3 US-08-746-559A-5 Sequence 4, Appli
41 139 9.2 516 3 US-08-746-559A-4 Sequence 2, Appli
42 1367 9.2 1367 2 US-08-249-687C-2 Sequence 2, Appli
43 139 9.2 1367 2 US-08-625-819-2 Sequence 2, Appli
44 139 9.2 1367 3 US-08-746-559A-2 Sequence 2, Appli
45 139 9.2 1367 4 US-08-864-641B-18 Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-08-284-941-2
; Sequence 2, Application US/08284941
; Patent No. 5863756
; GENERAL INFORMATION:
; APPLICANT: BARR, PHILIP J
; APPLICANT: KIEPER, MICHAEL C
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND
; TITLE OF INVENTION: PACE 4.1 GENE AND POLYPEPTIDES IN CELLS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESS: COOLEY GODWARD CASTRO HUDDLESON & TATUM
; STREET: FIVE PALO ALTO SQUARE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/284,941
; FILING DATE: 2 August 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: NEELLY PH.D., RICHARD L.
; REGISTRATION NUMBER: 30092
; REFERENCE/DOCKET NUMBER: CHIR-009/01US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 843-5070
; TELEFAX: (415) 857-0663
; TELEX: 380816 COOLEY PA
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 969 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULAR TYPE: protein
; US-08-284-941-2

Query Match 11.5%; Score 173.5; DB 2; Length 969;
Best Local Similarity 28.8%; Pred. No. 1.3e-06;
Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;
Oy 18 YIGSONASRGRRMRMHNVSGQGGCATCSD--YNGCLCKPRLFPALERIGMKQIGV 75
Db 737 YFGDTAARRCR-----CHKGCETSSRAATCLSCR-RGFY-----HHQEMNT 779

QY 76 CLSCPSGYGTRYPDINKCTCKKADCDTCFNK-NFCTCKSGPYLHGLKLDNCPGLE 134
Db 780 CVTLCPAGFYADB--SQNKCLKHPSCCKVDEPEKCTVCKBGFSLARGSCIPDCBPGTY 837
QY 135 ANNHTMECVSIVHCEVSEWNPSPCTKKGKTG 167
Db 838 FDSLEIRGEGCH-----TCG 853

RESULT 2
US-08-447-642-2
; Sequence 2, Application US/08447642
; Patent No. 5989890
; GENERAL INFORMATION:
; APPLICANT: BARR, PHILIP J
; APPLICANT: KIEFER, MICHAEL C
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND
; TITLE OF INVENTION: PACE 4.1 GENE AND POLYPEPTIDES IN CELLS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
; STREET: FIVE PALO ALTO SQUARE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,642
; FILING DATE: 23-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/284,941
; FILING DATE: 2 August 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: NEELEY PH.D., RICHARD L.
; REGISTRATION NUMBER: 30092
; REFERENCE/DOCKET NUMBER: CHIR-009/01US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 843-5070
; TELEFAX: (415) 857-0663
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 969 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-447-642-2

Query Match 11.5%; Score 173.5; DB 2; Length 969;
Best Local Similarity 28.8%; Pred. No. 1.3e-06;
Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;
QY 18 YIGSQNASRRQRMRHPNVSQGGGATCSD--YNGCLSKPRLPALERIGMKQIGV 75
Db 737 YFGDTAARRCR-----CHKGCTCSSRAATQCLSCR-RGPFY-----HHQEMNT 779
QY 76 CLSCPSGYGTRYPDINKCTCKKADCDTCFNK-NFCTCKSGPYLHGLKLDNCPGLE 134
Db 780 CVTLCPAGFYADB--SQNKCLKHPSCCKVDEPEKCTVCKBGFSLARGSCIPDCBPGTY 837
QY 135 ANNHTMECVSIVHCEVSEWNPSPCTKKGKTG 167
Db 838 FDSLEIRGEGCH-----TCG 853

; Sequence 2, Application US/09236503
; Patent No. 6277590
; GENERAL INFORMATION:
; APPLICANT: BARR, PHILIP J
; APPLICANT: KIEFER, MICHAEL C
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND 4.1 Gene and
; TITLE OF INVENTION: Polypeptides in Cells
; FILE REFERENCE: CHIR-009/04US
; CURRENT APPLICATION NUMBER: US/09/236,503
; CURRENT FILING DATE: 1999-01-25
; EARLIER APPLICATION NUMBER: 08/447,642
; EARLIER FILING DATE: 1995-05-23
; EARLIER APPLICATION NUMBER: 08/284,941
; EARLIER FILING DATE: 1994-08-02
; EARLIER APPLICATION NUMBER: 07/848,629
; EARLIER FILING DATE: 1992-03-09
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 969
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-236-503-2

Query Match 11.5%; Score 173.5; DB 3; Length 969;
Best Local Similarity 28.8%; Pred. No. 1.3e-06;
Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;
QY 18 YIGSQNASRRQRMRHPNVSQGGGATCSD--YNGCLSKPRLPALERIGMKQIGV 75
Db 737 YFGDTAARRCR-----CHKGCTCSSRAATQCLSCR-RGPFY-----HHQEMNT 779
QY 76 CLSCPSGYGTRYPDINKCTCKKADCDTCFNK-NFCTCKSGPYLHGLKLDNCPGLE 134
Db 780 CVTLCPAGFYADB--SQNKCLKHPSCCKVDEPEKCTVCKBGFSLARGSCIPDCBPGTY 837
QY 135 ANNHTMECVSIVHCEVSEWNPSPCTKKGKTG 167
Db 838 FDSLEIRGEGCH-----TCG 853

RESULT 4
PCT-US93-02147A-2
; Sequence 2, Application PC/TUS9302147A
; GENERAL INFORMATION:
; APPLICANT: BARR, PHILIP J
; APPLICANT: KIEFER, MICHAEL C
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND
; TITLE OF INVENTION: PACE 4.1 GENE AND POLYPEPTIDES IN CELLS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
; STREET: FIVE PALO ALTO SQUARE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/02147A
; FILING DATE: 19930309
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,629
; FILING DATE: 09-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: NEELEY PH.D., RICHARD L.
; REGISTRATION NUMBER: 30092
; REFERENCE/DOCKET NUMBER: CHIR-009/00US

```
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-7622
; TELEFAX: (415) 857-0663
; TELEX: 380816 COOLLEY PA
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 969 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-02147A-2

Query Match 11.5%; Score 173.5; DB 5; Length 969;
Best Local Similarity 28.8%; Pred. No. 1.3e-06;
Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;

QY 18 YIGSNASRGRRQRMRHNPVSGCGGCGCATCSD--YNGCLCKPLPFALEIRGMKQIGV 75
Db 737 YFGDTAARCR-----CHKGCETCSRAATQCLSCR-RGPF-----HHQEMNT 779

QY 76 CLSSCPGSGYGTTRYPDINKCTCKADCTCFNK-NFTYCKSGFYHLGKCLDNCPEGLE 134
Db 780 CVTLCPAGFYADE--SQKNCLCHPSCKKCVDEPEKCTVCKGFGSLARGSCIPCEPGTY 837

QY 135 ANNHTECVSVHCEVSEWNPSPCTKKGTGCG 167
Db 838 FDSILIRGCBCH-----TCG 853

RESULT 5
US-08-525-940-23
; Sequence 23, Application US/08525940
; Patent No. 5866351
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; APPLICANT: Miranda, Luis R.
; APPLICANT: Wolf, Joseph R.
; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES
; TITLE OF INVENTION: ENCODING SAID PROTEASES
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,940
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/368,852
; FILING DATE: 01-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/088,322
; FILING DATE: 07-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2848-11-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 799 amino acids
; TYPE: amino acid

;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-7622
; TELEFAX: (415) 857-0663
; TELEX: 380816 COOLLEY PA
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 969 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-525-940-23

Query Match 10.7%; Score 162; DB 2; Length 799;
Best Local Similarity 22.3%; Pred. No. 9.8e-06;
Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;

QY 20 GSONASRGRRQRMRHNPVSGCGGCGCATCSDYNG--CLSKPLPFALEIRGMKQIGVCL 77
Db 563 GHYHADK-KRCKKCAPN-----CESCFGSHGDDQMSCKGYFL-----NEETNSCV 607

QY 78 SSCPSG-YGTTRYPDINKCTCKADCTCFNK-NFTYCKSGFYHLGKCLDNCPEGLEAN 136
Db 608 THCPDGSQDTYK--NLCKKSENCKYCTEFTNCTECDGLSLOGSRCSVSCDGEYFN 664

QY 137 NHTME-----CVS-----IVHCEVSEW-----NPWSPCTKK 162
Db 665 GQDCQPCHRFCATCAGAGAGCGINCTEGYFMEBGRVCVQSCSISYFPHSSENGYKSCCKC 724

QY 163 GHTC-----GKRGTE-----TRVRELIQHPSAKGNLCPTTNETRKTCTV 201
Db 725 DISLTCNGPGPKNCTCSPSGYLLDGMCMGAICDKATSESWAEGGFCMLVKKNILC-- 782

QY 202 QRKKQCK 208
Db 783 QRKVLQQ 789

RESULT 6
US-08-976-838-23
; Sequence 23, Application US/08976838
; Patent No. 5981259
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,838
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2848-11-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 799 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-976-838-23

Query Match 10.7%; Score 162; DB 2; Length 799;
Best Local Similarity 22.3%; Pred. No. 9.8e-06;
Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;
```


QY 137 NHTME-----CVS-----IVHCEVSEW-----NPWSPTCK 162
 Db 747 QDCQPCHRFCATCAGACAGCINTEGYFMEDGRCVQSCSISYFDSSENGYKSKKC 806
 QY 163 GKTG-----GFKRGTE-----TRVREIIQHPSAKGNLCPPPTNETRKTCTV 201
 Db 807 DISCLTCNCGPGFKNCTCPSGYLLDLGCMQGAICKDATEESWABGGFCMLVKKNLNC-- 864
 QY 202 QKKCKOK 208
 Db 865 QKKVLOQ 871

RESULT 9

US-08-525-940-18
 ; Sequence 18, Application US/08525940
 ; Patent No. 586351
 ; GENERAL INFORMATION:
 ; APPLICANT: Franzusoff, Alex
 ; APPLICANT: Miranda, Luis R.
 ; APPLICANT: Wolf, Joseph R.
 ; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES
 ; TITLE OF INVENTION: ENCODING SAID PROTEASES
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sheridan Ross & McIntosh
 ; STREET: 1700 Lincoln Street, Suite 3500
 ; CITY: Denver
 ; STATE: Colorado
 ; COUNTRY: U.S.A.
 ; ZIP: 80203
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/525,940
 ; FILING DATE:
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/368,852
 ; FILING DATE: 01-JAN-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/088,322
 ; FILING DATE: 07-JUL-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Connell, Gary J.
 ; REGISTRATION NUMBER: 32,020
 ; REFERENCE/DOCKET NUMBER: 2848-11-C1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (303) 863-9700
 ; TELEFAX: (303) 863-0223
 ; INFORMATION FOR SEQ ID NO: 18:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 915 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-525-940-18

Query Match 10.7%; Score 162; DB 2; Length 915;
 Best Local Similarity 22.3%; Pred. No. 1.1e-05;
 Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;
 QY 20 GSNASRGRRORRHPPNVSCQCGGCATCSDYNG--CLSCKPRLFFALERIGMKQIGVCL 77
 Db 679 GHYHADK-KRCKCAPN-----CESCFGSHGDDQMSCKYGYFL-----NEETNSCV 723
 QY 78 SSCPSG--YGTTRYPDINKTKKADCDTCFNKNFCTCKSGFYHLGKCLDNCPEGLAN 136
 Db 724 THCPDGSVQDTKK---NLCKRCSENCCKTCTBFHNTCTEGRDGLSLQGSRCVSCEGGRYFN 780

QY 137 NHTME-----CVS-----IVHCEVSEW-----NPWSPTCK 162
 Db 781 QDCQPCHRFCATCAGACAGCINTEGYFMEDGRCVQSCSISYFDSSENGYKSKKC 840
 QY 163 GKTG-----GFKRGTE-----TRVREIIQHPSAKGNLCPPPTNETRKTCTV 201
 Db 841 DISCLTCNCGPGFKNCTCPSGYLLDLGCMQGAICKDATEESWABGGFCMLVKKNLNC-- 898
 QY 202 QKKCKOK 208
 Db 899 QKKVLOQ 905

RESULT 10
 US-08-976-838-18
 ; Sequence 18, Application US/08976838
 ; Patent No. 5981259
 ; GENERAL INFORMATION:
 ; APPLICANT: Franzusoff, Alex
 ; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID
 ; TITLE OF INVENTION: MOLECULES
 ; NUMBER OF SEQUENCES: 31
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sheridan Ross P.C.
 ; STREET: 1700 Lincoln St., Suite 3500
 ; CITY: Denver
 ; STATE: Colorado
 ; COUNTRY: U.S.A.
 ; ZIP: 80203
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/976,838
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Connell, Gary J.
 ; REGISTRATION NUMBER: 32,020
 ; REFERENCE/DOCKET NUMBER: 2848-11-C2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (303) 863-9700
 ; TELEFAX: (303) 863-0223
 ; INFORMATION FOR SEQ ID NO: 18:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 915 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-976-838-18

Query Match 10.7%; Score 162; DB 2; Length 915;
 Best Local Similarity 22.3%; Pred. No. 1.1e-05;
 Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;
 QY 20 GSNASRGRRORRHPPNVSCQCGGCATCSDYNG--CLSCKPRLFFALERIGMKQIGVCL 77
 Db 679 GHYHADK-KRCKCAPN-----CESCFGSHGDDQMSCKYGYFL-----NEETNSCV 723
 QY 78 SSCPSG--YGTTRYPDINKTKKADCDTCFNKNFCTCKSGFYHLGKCLDNCPEGLAN 136
 Db 724 THCPDGSVQDTKK---NLCKRCSENCCKTCTBFHNTCTEGRDGLSLQGSRCVSCEGGRYFN 780
 QY 137 NHTME-----CVS-----IVHCEVSEW-----NPWSPTCK 162
 Db 781 QDCQPCHRFCATCAGACAGCINTEGYFMEDGRCVQSCSISYFDSSENGYKSKKC 840
 QY 163 GKTG-----GFKRGTE-----TRVREIIQHPSAKGNLCPPPTNETRKTCTV 201
 Db 841 DISCLTCNCGPGFKNCTCPSGYLLDLGCMQGAICKDATEESWABGGFCMLVKKNLNC-- 898

```
QY 202 QKCKCK 208
Db 899 QKVLQ 905

RESULT 11
US-09-214-555B-2
; Sequence 2, Application US/09214555B
; Patent No. 6380171
; GENERAL INFORMATION:
; APPLICANT: INSTITUT DE RECHERCHE CLINIQUE DE MONTREAL
; TITLE OF INVENTION: PRO-PROTEIN CONVERTING ENZYME
; FILE REFERENCE: PRO-PROTEIN CONVER ENZ
; CURRENT APPLICATION NUMBER: US/09/214,555B
; CURRENT FILING DATE: 1999-01-04
; PRIOR APPLICATION NUMBER: 60/021,008
; PRIOR FILING DATE: 1996-07-26
; PRIOR APPLICATION NUMBER: 2,203,745
; PRIOR FILING DATE: 1997-04-25
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 2
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-555B-2

Query Match 10.7%; Score 162; DB 4; Length 915;
Best Local Similarity 22.3%; Pred. No. 1.1e-05;
Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;

QY 20 GSONASRGRRQRHMHVNSQGGCGCATCSDYNG--CLSCPKPLFFALERIGMKQIGVCL 77
Db 679 GHYHADK-KRCRKAPE-----CESCFGSHGDQMSCKYGYFL-----NEETNSCV 723

QY 78 SSCPSG-YGYTRYPDINKCTKADCDTCFNKNFCTKCKSGFYHLGKCLDNCPEGLEAN 136
Db 724 THCPDSGYQDTTK--NLCKRCSNCKYCTEFNCTCECDGLSLQGRCSVSCDGRYFN 780

QY 137 NHTME-----CVS-----IVHCEVSEW-----NPMSPCTKK 162
Db 781 GQDCQPCHRFCATCAGAGDGCINCTEGYFMEDGRCVQSCSIYYPDHSSENGYKSKCK 840

QY 163 GKTC-----GPKRGTE-----TRVREIIQHPSAKGNLCPPPTNTRKCTV 201
Db 841 DISCLTCNGPGFNCTSCPSGYLLDGMQMGAIKDAEESWABGGFCMLVKKNLCL-- 898

QY 202 QKCKCK 208
Db 899 QKVLQ 905

RESULT 12
US-09-214-555B-7
; Sequence 7, Application US/09214555B
; Patent No. 6380171
; GENERAL INFORMATION:
; APPLICANT: INSTITUT DE RECHERCHE CLINIQUE DE MONTREAL
; TITLE OF INVENTION: PRO-PROTEIN CONVERTING ENZYME
; FILE REFERENCE: PRO-PROTEIN CONVER ENZ
; CURRENT APPLICATION NUMBER: US/09/214,555B
; CURRENT FILING DATE: 1999-01-04
; PRIOR APPLICATION NUMBER: 60/021,008
; PRIOR FILING DATE: 1996-07-26
; PRIOR APPLICATION NUMBER: 2,203,745
; PRIOR FILING DATE: 1997-04-25
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 7
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-555B-7

Query Match 10.7%; Score 161; DB 1; Length 288;
Best Local Similarity 28.2%; Pred. No. 3.8e-06;
Matches 46; Conservative 25; Mismatches 52; Indels 40; Gaps 11;

QY 35 PNVSQ-GCQG-GCATCSDYNGCLSKCRPLFFALERIGMKQIGVCLSLSCPSGYYGTRYDPI 92
```


Db 13 PECSVGCDCGPDHCND---CL-----HYVYKLGK-----NNTRICVSSCPPGHY---HADK 58
Qy 93 NKCTKADCTCFNK--NFCCKSGFYLH--LGKCLDNCPEGLAANNHTBECVSIYHC 148
Db 59 KRCRCAPNCSCFSGHGDQCMCKYGYFLNEETNSCVTHCPDGSYDPTKKNLC----- 112
Qy 149 EVSEWNPSPCTKKGKTC-GFKRGTTETRVREIIQHPSAKGNLC 190
Db 113 -----RKSENKKIUTETFAKTECR-----DGLSLQGSRC 142

RESULT 14

US-09-907-794A-4

; Sequence 4, Application US/09907794A
; Patent No. 663468
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas P.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,794A
; PRIOR FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095

; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 4
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-907-794A-4

Query Match 10.5%; Score 158.5; DB 4; Length 379;

Best Local Similarity 24.9%; Pred. No. 8.5e-06;
Matches 60; Conservative 26; Mismatches 72; Indels 83; Gaps 16;

Qy 41 CQGCAT---CSDVNGC-----LSCKRLPFALERIGMKQIGVCLSS---C 80
Db 182 CPGCGNGGFCNERRICEPCDPGPHCEKALCTPCMG-----GLCVTPGFCIC 232
Qy 81 PSYYGTRYPDINKTKC-KADCD-TCFNKNFTCKKSGFYHLGKCLDNCPEGLAANNH 138
Db 233 PPGFYG-----VNCCKANCSSTCFNGGTC-----FY--PGKCI--CPPGLEGE-- 271
Qy 139 TMECVSIHVHCEVSEWNPSPCTKKGKTCGFKRGTTETRVREIIQHPSAKGNLCPP----- 192
Db 272 -----QCHSKCP--QPCRNGGKCTG-----KSKCKSGKYQGLCLCKPVCPCGAHG 318
Qy 193 -TNETRKTQVQKKCKGKRGKGRKRRK-----KPNKGSKEAIPDSKLSLESSKEIP 245
Db 319 TCHEPNKC-----CQCEGWHGRHCNKRYEASLIHALRPAQALRQHTTFLSKAEERDPP 373
Qy 246 E 246
Db 374 E 374

RESULT 15

US-09-905-125A-4

; Sequence 4, Application US/09905125A
; Patent No. 6664376
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas P.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905,125A

Wed Jun 30 07:55:47 2004

```
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 4
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-905-125A-4

Query Match      10.5%; Score 158.5; DB 4; Length 379;
Best Local Similarity 24.9%; Pred. No. 8.5e-06;
Matches 60; Conservative 26; Mismatches 72; Indels 83; Gaps 16;

QY      41  CQGCAT---CSDYNGC-----LSCKPLFFALERIGMKQIGVCLSS-----C 80
DB      182  CPGGCRNGGFCNERRICEPCDGFHGHPCERKALCTPRCNG-----GLCVTPGFCIC 232

QY      81  PSGYGTRYDPINKCTKC-KADCD-TCFNKNFCTKCKSGFYHLGKLDNCPGLEANNH 138
DB      233  PPGFYG-----VNCDKANCSITCFNGGTC-----PY--FGKCI--CPPGLEGE-- 271

QY      139  TMECVSIHCRVSEMNFWSPCTKKGKTCGPKRGKTETRVREIIQHPSAKGNLCP----- 192
DB      272  -----QCISKCP--QPCRNGKCIQ---KSKCKSKGYQDLSKPVCFPGCGGANG 318

QY      193  -TNETRKTQVORKKCKGKGRKRRK-----KFNKGESKKAIPDSKLSSESKHIP 245
DB      319  TCHFPKNC-----CQSGWGHGRHCNRYEASLIHALRPAQAQLRQHTPFLAKAERRDPP 373

QY      246  E 246
DB      374  E 374
```

Search completed: June 29, 2004, 17:04:02
Job time : 15.1942 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 17:02:26 ; Search time 36.6408 Seconds

(without alignments)
2098.641 Million cell updates/sec

Title: US-09-894-912A-34

Perfect score: 1510

Sequence: 1 MHLRLISWLFILNFMFYIG.....QQKKRKVDKQKSVSVTVH 272

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1166195 seqs, 282705291 residues

Total number of hits satisfying chosen parameters: 1166195

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2_6/ptodata/1/pubpa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1510	100.0	272	10	US-09-894-912A-10
2	1510	100.0	272	10	US-09-894-912A-34
3	1510	100.0	272	14	US-10-125-852-23
4	1505	99.7	273	10	US-09-894-912A-13
5	1477	97.8	265	10	US-09-894-912A-26
6	1477	97.8	292	9	US-09-745-763-166
7	1477	97.8	292	10	US-09-894-912A-48
8	1477	97.8	292	14	US-10-125-852-25
9	1400	92.7	251	10	US-09-894-912A-16
10	1315.5	87.1	279	10	US-09-894-912A-32
11	1280	84.8	225	14	US-10-185-770-4
12	1218	80.7	239	12	US-10-087-192-1782
13	1183	78.3	239	15	US-10-094-886-172
14	1111	73.6	195	15	US-10-094-886-176
15	1107	73.3	195	15	US-10-094-886-174

16	981	65.0	180	12	US-10-087-192-1779	Sequence 1779, Ap
17	903	59.8	160	10	US-09-894-912A-14	Sequence 14, Appl
18	903	59.8	160	12	US-10-276-774-1744	Sequence 1744, Ap
19	656	43.4	263	14	US-10-125-852-18	Sequence 18, Appl
20	648.5	42.9	243	14	US-10-125-852-21	Sequence 21, Appl
21	644	42.6	265	14	US-10-125-852-24	Sequence 24, Appl
22	638	42.3	229	10	US-09-894-912A-25	Sequence 25, Appl
23	626.5	41.5	236	15	US-10-108-260A-4829	Sequence 4829, Ap
24	589.5	39.0	243	14	US-10-185-770-2	Sequence 2, Appli
25	582.5	38.6	243	14	US-10-125-852-13	Sequence 13, Appl
26	582.5	38.6	243	16	US-10-467-042-12	Sequence 12, Appl
27	578	38.3	250	14	US-10-125-852-3	Sequence 3, Appli
28	545	36.1	222	14	US-10-125-852-15	Sequence 15, Appl
29	540.5	35.8	229	14	US-10-125-852-6	Sequence 6, Appli
30	402.5	26.7	190	12	US-10-087-192-873	Sequence 873, App
31	365.5	24.2	161	12	US-10-087-192-876	Sequence 876, App
32	276	18.3	46	10	US-09-894-912A-18	Sequence 18, Appl
33	239.5	15.9	131	14	US-10-125-852-9	Sequence 9, Appli
34	221	14.6	42	10	US-09-894-912A-22	Sequence 22, Appl
35	213	14.1	37	10	US-09-894-912A-20	Sequence 20, Appl
36	202	13.4	110	14	US-10-125-852-11	Sequence 11, Appl
37	186.5	12.4	1548	14	US-10-180-903-2	Sequence 2, Appli
38	177.5	11.8	43	14	US-10-125-852-7	Sequence 7, Appli
39	173.5	11.5	969	10	US-09-961-403-6	Sequence 6, Appli
40	160.5	10.6	337	12	US-10-357-820-20	Sequence 20, Appl
41	160.5	10.6	343	12	US-10-357-820-8	Sequence 8, Appli
42	160.5	10.6	365	12	US-10-357-820-4	Sequence 4, Appli
43	160.5	10.6	365	12	US-10-357-820-6	Sequence 6, Appli
44	160.5	10.6	373	12	US-10-357-820-10	Sequence 10, Appl
45	160.5	10.6	479	9	US-09-764-898-221	Sequence 221, App

ALIGNMENTS

RESULT 1

US-09-894-912A-10
; Sequence 10, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; TITLE OF INVENTION: FACTOR-LIKE POLYPEPTIDES AND POLYNUCLEOTIDES
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; CURRENT FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-894-912A-10

Query Match 100.0%; Score 1510; DB 10; Length 272;
Best Local Similarity 100.0%; Pred. No. 5.2e-115;
Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MHLRLISWLFILNFMFYIGSNASRGRRQRMRHNPVSGGCGCATCDYNGCLSCRP	60
Db	1	MHLRLISWLFILNFMFYIGSNASRGRRQRMRHNPVSGGCGCATCDYNGCLSCRP	60
Qy	61	LPFALRIGMKQIGVGLSSCPGGYGYTRYPDINKTKCKADCDTCFNFKTKKSGFY	120

```
Db 61 LFFALERIGMKQIGVCLSSCPGGYGYTRYPDINKTKCADCTCFNKNFCTKCKSGFYL 120
Qy 121 HLKCLDNCPEGLANNTMECVSIHVCEVSEWNPSPCTKKGTCGFRGTETRVREII 180
Db 121 HLKCLDNCPEGLANNTMECVSIHVCEVSEWNPSPCTKKGTCGFRGTETRVREII 180
Qy 181 QHPSAKGNLCPPTNETRKTCTVQRKCKGKGRKRRKPKPKNGSKKAIPOKSKLES 240
Db 181 QHPSAKGNLCPPTNETRKTCTVQRKCKGKGRKRRKPKPKNGSKKAIPOKSKLES 240
Qy 241 SKIPEORENKQKKRKVKQDKKSKSVSVTVH 272
Db 241 SKIPEORENKQKKRKVKQDKKSKSVSVTVH 272
```

RESULT 2

```
US-09-894-912A-34
; Sequence 34, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; PRIOR FILING DATE: 2002-05-10
; PRIOR FILING DATE: 2001-04-05
; PRIOR FILING DATE: 2001-02-05
; PRIOR FILING DATE: 2001-02-05
; PRIOR FILING DATE: 2000-06-28
; PRIOR FILING DATE: 2000-06-28
; PRIOR FILING DATE: 2001-01-09
; PRIOR FILING DATE: 2001-01-09
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-894-912A-34
```

```
Query Match 100.0%; Score 1510; DB 10; Length 272;
Best Local Similarity 100.0%; Pred. No. 5.2e-115;
Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHLRLISWLFILNFMVEYIGSNASRGRORRMRHPNVSQGGCGCATCSDYNGCLSKPR 60
Db 1 MHLRLISWLFILNFMVEYIGSNASRGRORRMRHPNVSQGGCGCATCSDYNGCLSKPR 60
Qy 61 LFFALERIGMKQIGVCLSSCPGGYGYTRYPDINKTKCADCTCFNKNFCTKCKSGFYL 120
Db 61 LFFALERIGMKQIGVCLSSCPGGYGYTRYPDINKTKCADCTCFNKNFCTKCKSGFYL 120
Qy 121 HLKCLDNCPEGLANNTMECVSIHVCEVSEWNPSPCTKKGTCGFRGTETRVREII 180
Db 121 HLKCLDNCPEGLANNTMECVSIHVCEVSEWNPSPCTKKGTCGFRGTETRVREII 180
Qy 181 QHPSAKGNLCPPTNETRKTCTVQRKCKGKGRKRRKPKPKNGSKKAIPOKSKLES 240
Db 181 QHPSAKGNLCPPTNETRKTCTVQRKCKGKGRKRRKPKPKNGSKKAIPOKSKLES 240
Qy 241 SKIPEORENKQKKRKVKQDKKSKSVSVTVH 272
Db 241 SKIPEORENKQKKRKVKQDKKSKSVSVTVH 272
```

RESULT 3

```
US-10-125-852-23
; Sequence 23, Application US/10125852
```

```
; Publication No. US20030032034A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH FACTOR-
; FILE REFERENCE: HYS-43A
; CURRENT APPLICATION NUMBER: US/10/125,852
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: US 60/316,368
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: US 09/799,451
; PRIOR FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-125-852-23
```

```
Query Match 100.0%; Score 1510; DB 14; Length 272;
Best Local Similarity 100.0%; Pred. No. 5.2e-115;
Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHLRLISWLFILNFMVEYIGSNASRGRORRMRHPNVSQGGCGCATCSDYNGCLSKPR 60
Db 1 MHLRLISWLFILNFMVEYIGSNASRGRORRMRHPNVSQGGCGCATCSDYNGCLSKPR 60
Qy 61 LFFALERIGMKQIGVCLSSCPGGYGYTRYPDINKTKCADCTCFNKNFCTKCKSGFYL 120
Db 61 LFFALERIGMKQIGVCLSSCPGGYGYTRYPDINKTKCADCTCFNKNFCTKCKSGFYL 120
Qy 121 HLKCLDNCPEGLANNTMECVSIHVCEVSEWNPSPCTKKGTCGFRGTETRVREII 180
Db 121 HLKCLDNCPEGLANNTMECVSIHVCEVSEWNPSPCTKKGTCGFRGTETRVREII 180
Qy 181 QHPSAKGNLCPPTNETRKTCTVQRKCKGKGRKRRKPKPKNGSKKAIPOKSKLES 240
Db 181 QHPSAKGNLCPPTNETRKTCTVQRKCKGKGRKRRKPKPKNGSKKAIPOKSKLES 240
Qy 241 SKIPEORENKQKKRKVKQDKKSKSVSVTVH 272
Db 241 SKIPEORENKQKKRKVKQDKKSKSVSVTVH 272
```

RESULT 4

```
US-09-894-912A-13
; Sequence 13, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; CURRENT FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-894-912A-13
```

```

Db      181  QHPSAGNLCPPTNETKCTVQRKKCKQGRGKKGRKRKPKNGKSEKALPDSKLS 240
QY      241  SKEIPEQRENKQOQKKRKVQDKKS 265
        |||||
Db      241  SKEIPEQRENKQOQKKRKVQDKKS 265
        |||||

RESULT 6
US-09-745-763-166
; Sequence 166, Application US/09745763
; Patent No. US20020065394A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; McCoy, John M.
; Lavallie, Edward R.
; Collins-Racie, Lisa A.
; Evans, Cheryl
; Merberg, David
; Treacy, Maurice
; Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; ENCODING THEM
; NUMBER OF SEQUENCES: 219
; CORRESPONDENCE ADDRESS:
;

```

CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/745,763
FILING DATE: 18-Jun-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,323
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 166:
SEQUENCE CHARACTERISTICS:
LENGTH: 292 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 166:
US-09-745-763-166

	Query Match	97.8%	Score 1477;	DB 9;	Length 292;	
	Best Local Similarity	100.0%;	Posed. No. 2.7e-112;			
	Matches 265;	Conservative	0;	Mismatches	0;	Gaps 0
Qy	1	MHLRLISWLPIILNFMFYIYISQNASRGRQRORMPNVSQGGCGCATCSDYNGCLSCSKRP				60
Db	1	MHLRLISWLPIILNFMFYIYISQNASRGRQRORMPNVSQGGCGCATCSDYNGCLSCSKRP				60
Qy	61	LFPALERIGMKQIGVCLSSCPSGYGYTRYDPDINCKTCKADCDTCFNKNFCTCKSGGYL				120
Db	61	LFPALERIGMKQIGVCLSSCPSGYGYTRYDPDINCKTCKADCDTCFNKNFCTCKSGGYL				120
Qy	121	HLGKCLDNCPEGLNANHNTWECVSIHVCEVSEWNPSPCTKKGKTGCFKRGGTETVREII				180
Db	121	HLGKCLDNCPEGLNANHNTWECVSIHVCEVSEWNPSPCTKKGKTGCFKRGGTETVREII				180
Qy	181	QHPESAKNLCPPNNTETKCTVQRKKCKGGRGKGRBRKCKPKNGSKSEALPDKSKSLK				240

Db 181 QHPSAKGNLCPTNETRKTCTVQRKCKQKGRGKGRKRRKRRKPKNKGSKEAIPDSKSLES 240
QY 241 SKEIPEORENKQKKRKKVQDKOKS 265
Db 241 SKEIPEORENKQKKRKKVQDKOKS 265

RESULT 7

US-09-894-912A-48
; Sequence 48, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; FILE REFERENCE: 28110/37260A
; CURRENT FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: US/09/894,912A
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 48
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-894-912A-48

Query Match 97.8%; Score 1477; DB 10; Length 292;
Best Local Similarity 100.0%; Pred. No. 2.7e-112; Indels 0; Gaps 0;
Matches 265; Conservative 0; Mismatches 0; Mismatches 0; Mismatches 0; Mismatches 0;
QY 1 MHLRLISWLFILNFMFYIGSNASRGRQRMRHHPNVSQCGGCATCSDYNGCLSCPKR 60
Db 1 MHLRLISWLFILNFMFYIGSNASRGRQRMRHHPNVSQCGGCATCSDYNGCLSCPKR 60
QY 61 LPPALERIGMKQIGVCLSSCPSGYTTRYPDINKCTCKADCDTCFKNKFCCKSGPYL 120
Db 61 LPPALERIGMKQIGVCLSSCPSGYTTRYPDINKCTCKADCDTCFKNKFCCKSGPYL 120
QY 121 HLGKCLDNCPEGLANNHTMECVSIHVCEVSEWNPSPCTKKGTCGFKRGTTETRVREII 180
Db 121 HLGKCLDNCPEGLANNHTMECVSIHVCEVSEWNPSPCTKKGTCGFKRGTTETRVREII 180
QY 181 QHPSAKGNLCPTNETRKTCTVQRKCKQKGRGKGRKRRKRRKPKNKGSKEAIPDSKSLES 240
Db 181 QHPSAKGNLCPTNETRKTCTVQRKCKQKGRGKGRKRRKRRKPKNKGSKEAIPDSKSLES 240
QY 241 SKEIPEORENKQKKRKKVQDKOKS 265
Db 241 SKEIPEORENKQKKRKKVQDKOKS 265

RESULT 8

US-10-125-852-25
; Sequence 25, Application US/10125852
; Publication No. US20030032034A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH FACTOR-1
; FILE REFERENCE: HYS-43A
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: US 60/316,368

; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: US 09/799,451
; PRIOR FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-125-852-25

Query Match 97.8%; Score 1477; DB 14; Length 292;
Best Local Similarity 100.0%; Pred. No. 2.7e-112; Indels 0; Gaps 0;
Matches 265; Conservative 0; Mismatches 0; Mismatches 0; Mismatches 0; Mismatches 0;
QY 1 MHLRLISWLFILNFMFYIGSNASRGRQRMRHHPNVSQCGGCATCSDYNGCLSCPKR 60
Db 1 MHLRLISWLFILNFMFYIGSNASRGRQRMRHHPNVSQCGGCATCSDYNGCLSCPKR 60
QY 61 LPPALERIGMKQIGVCLSSCPSGYTTRYPDINKCTCKADCDTCFKNKFCCKSGPYL 120
Db 61 LPPALERIGMKQIGVCLSSCPSGYTTRYPDINKCTCKADCDTCFKNKFCCKSGPYL 120
QY 121 HLGKCLDNCPEGLANNHTMECVSIHVCEVSEWNPSPCTKKGTCGFKRGTTETRVREII 180
Db 121 HLGKCLDNCPEGLANNHTMECVSIHVCEVSEWNPSPCTKKGTCGFKRGTTETRVREII 180
QY 181 QHPSAKGNLCPTNETRKTCTVQRKCKQKGRGKGRKRRKRRKPKNKGSKEAIPDSKSLES 240
Db 181 QHPSAKGNLCPTNETRKTCTVQRKCKQKGRGKGRKRRKRRKPKNKGSKEAIPDSKSLES 240
QY 241 SKEIPEORENKQKKRKKVQDKOKS 265
Db 241 SKEIPEORENKQKKRKKVQDKOKS 265

RESULT 9

US-09-894-912A-16
; Sequence 16, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; FILE REFERENCE: 28110/37260A
; CURRENT FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-894-912A-16

Query Match 92.7%; Score 1400; DB 10; Length 251;
Best Local Similarity 100.0%; Pred. No. 4.3e-106; Indels 0; Gaps 0;
Matches 251; Conservative 0; Mismatches 0; Mismatches 0; Mismatches 0; Mismatches 0;
QY 22 QNASRGRQRMRHHPNVSQCGGCATCSDYNGCLSCPKRPFALERIGMKQIGVCLSSCP 81
Db 1 QNASRGRQRMRHHPNVSQCGGCATCSDYNGCLSCPKRPFALERIGMKQIGVCLSSCP 60

QY 82 SGYYGTRYPDINKCTCKKADCDTCFKNKPFCTCKSGFYHLGKCLDNCBGLNANNHTWE 141
DB 61 SGYYGTRYPDINKCTCKKADCDTCFKNKPFCTCKSGFYHLGKCLDNCBGLNANNHTWE 120
QY 142 CVSIVHCVSEVSNWSPCTCKGKTGCKRGTRVREIIQHPSAKGNLCPPPTNETRKTCTV 201
DB 121 CVSIVHCVSEVSNWSPCTCKGKTGCKRGTRVREIIQHPSAKGNLCPPPTNETRKTCTV 180
QY 202 ORKKCKGKRGKGRKRRKKNKESKEAIPDSKLSLESSKEIPEORENKKOQKKRVOD 261
DB 181 ORKKCKGKRGKGRKRRKKNKESKEAIPDSKLSLESSKEIPEORENKKOQKKRVOD 240
QY 262 KOKSVSVSTVH 272
DB 241 KOKSVSVSTVH 251

RESULT 10

US-09-894-912A-32
; Sequence 32, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; CURRENT FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 32
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-894-912A-32

Query Match 87.1%; Score 1315.5; DB 10; Length 279;
Best Local Similarity 87.1%; Pred. No. 3.7e-99;
Matches 243; Conservative 11; Mismatches 18; Indels 7; Gaps 4;
QY 1 MHLRLISWLFILNFMFYIGSNASRRGRQRMRHNPVSGCGGCATCSDYNGCLSCKPR 60
DB 1 MHLRLISWLFILNFMFYIGSNASRRGRQRMRHNPVSGCGGCATCSDYNGCLSCKPR 60
QY 61 LFFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKCTCKKADCDTCFKNKPFCTCKSGFY 120
DB 61 LFFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKCTCKKADCDTCFKNKPFCTCKSGFY 120
QY 121 HLGKCLDNCBGLNANNHTMCVSIHCVSEVSNWSPCTCKGKTGCKRGTRVREII 180
DB 121 HLGKCLDNCBGLNANNHTMCVSIHCVSEVSNWSPCTCKGKTGCKRGTRVREII 180
QY 181 QHPSA--KGNLCPPPTNETRKTCTVORKKCKGKGRKRRKKNKESKE--AIPDSK 236
DB 181 QHPSAKGRGNLCPPPTSETRTCTVORKKCKGKGRKRRKKNKESKE--AIPDSK 240
QY 237 SLESSKEIPEORENKK--OQKKRVODK--OKSVSVSTVH 272
DB 241 GLESSIETPDQENKRRK--OQKKRRARDK--OKSVSVSTVH 279

RESULT 11

US-10-185-770-4

; Sequence 4, Application US/10185770
; Publication No. US20030022217A1
; GENERAL INFORMATION:
; APPLICANT: CECCARDI, Toni et al.
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; FILE REFERENCE: CL0001247
; CURRENT APPLICATION NUMBER: US/10/185,770
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/301,852
; PRIOR FILING DATE: 2001-07-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-185-770-4

Query Match 84.8%; Score 1280; DB 14; Length 225;
Best Local Similarity 100.0%; Pred. No. 2.2e-96;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHLRLISWLFILNFMFYIGSNASRRGRQRMRHNPVSGCGGCATCSDYNGCLSCKPR 60
DB 1 MHLRLISWLFILNFMFYIGSNASRRGRQRMRHNPVSGCGGCATCSDYNGCLSCKPR 60
QY 61 LFFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKCTCKKADCDTCFKNKPFCTCKSGFY 120
DB 61 LFFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKCTCKKADCDTCFKNKPFCTCKSGFY 120
QY 121 HLGKCLDNCBGLNANNHTMCVSIHCVSEVSNWSPCTCKGKTGCKRGTRVREII 180
DB 121 HLGKCLDNCBGLNANNHTMCVSIHCVSEVSNWSPCTCKGKTGCKRGTRVREII 180
QY 181 QHPSAKGNLCPPPTNETRKTCTVORKKCKGKGRKRRKKNK 225
DB 181 QHPSAKGNLCPPPTNETRKTCTVORKKCKGKGRKRRKKNK 225

RESULT 12

US-10-087-192-1782
; Sequence 1782, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1782
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-087-192-1782

Query Match 80.7%; Score 1218; DB 12; Length 239;
Best Local Similarity 97.3%; Pred. No. 2.7e-91;
Matches 215; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 1 MHLRLISWLFILNFMFYIGSNASRRGRQRMRHNPVSGCGGCATCSDYNGCLSCKPR 60
DB 18 MHLRLISWLFILNFMFYIGSNASRRGRQRMRHNPVSGCGGCATCSDYNGCLSCKPR 77


```
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/318,510
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/314,018
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/296,693
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/313,626
; PRIOR FILING DATE: 2001-08-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 298
; SOFTWARE: PatentIn 2.1
; SEQ ID NO 176
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-886-176

Query Match      73.6%; Score 1111; DB 15; Length 195;
Best Local Similarity 100.0%; Pred. No. 1.1e-82;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 GSNASRGRQRMRHPNVSQCGGCATCSDYNGCLSKPRLPFALERIGMKQIGVCLSS 79
Db 1 GSNASRGRQRMRHPNVSQCGGCATCSDYNGCLSKPRLPFALERIGMKQIGVCLSS 60

Qy 80 CPSSGYGTRYPDINKCKKADCDTCFNKNFCTCKSGFYHLGKCLDNCPEGLEANNHT 139
Db 61 CPSSGYGTRYPDINKCKKADCDTCFNKNFCTCKSGFYHLGKCLDNCPEGLEANNHT 120

Qy 140 MECVSIHVCEVSEWNPSPCTKKGKTCGKRGTTREIIOHPSAKGNLCPTNETRKC 199
Db 121 MECVSIHVCEVSEWNPSPCTKKGKTCGKRGTTREIIOHPSAKGNLCPTNETRKC 180

Qy 200 TVQRKKCKQKGERG 212
Db 181 TVQRKKCKQKGERG 193

RESULT 15
US-10-094-886-174
; Sequence 174, Application US/10094886
; Publication No. US20040002120A1
; GENERAL INFORMATION:
; APPLICANT: Kektuda, Ramesh
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Pattarajan, Meera
; APPLICANT: Burgess, Catherine
; APPLICANT: Vernet, Corine A.
; APPLICANT: Li, Li
; APPLICANT: Gorman, Linda
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Boldog, Ferenc
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shenoy, Suresh
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Taupier, Raymond J., Jr.
; APPLICANT: Miller, Charles
; APPLICANT: Casman, Stacie
; APPLICANT: Pena, Carol
; APPLICANT: Gangoli, Baha
; APPLICANT: Gusev, Vladimir
; APPLICANT: Smithson, Glennda
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Gerlach, Valerie
```

```
; APPLICANT: Pochart, Pascal
; APPLICANT: Fernandes, Elma
; APPLICANT: Shinkets, Richard
; APPLICANT: Rastelli, Luca
; APPLICANT: Spaderna, Steven
; APPLICANT: LaRoche, William
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-290 B
; CURRENT APPLICATION NUMBER: US/10/094,886
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/274,322
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/313,182
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/288,052
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/318,510
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/314,018
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/296,693
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/313,626
; PRIOR FILING DATE: 2001-08-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 298
; SOFTWARE: PatentIn 2.1
; SEQ ID NO 174
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-886-174

Query Match      73.3%; Score 1107; DB 15; Length 195;
Best Local Similarity 99.5%; Pred. No. 2.3e-82;
Matches 192; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 20 GSNASRGRQRMRHPNVSQCGGCATCSDYNGCLSKPRLPFALERIGMKQIGVCLSS 79
Db 1 GSNASRGRQRMRHPNVSQCGGCATCSDYNGCLSKPRLPFALERIGMKQIGVCLSS 60

Qy 80 CPSSGYGTRYPDINKCKKADCDTCFNKNFCTCKSGFYHLGKCLDNCPEGLEANNHT 139
Db 61 CPSSGYGTRYPDINKCKKADCDTCFNKNFCTCKSGFYHLGKCLDNCPEGLEANNHT 120

Qy 140 MECVSIHVCEVSEWNPSPCTKKGKTCGKRGTTREIIOHPSAKGNLCPTNETRKC 199
Db 121 MECVSIHVCEVSEWNPSPCTKKGKTCGKRGTTREIIOHPSAKGNLCPTNETRKC 180

Qy 200 TVQRKKCKQKGERG 212
Db 181 TVQRKKCKQKGERG 193
```

Search completed: June 29, 2004, 17:15:58
Job time : 37.6408 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 16:56:05 ; Search time 13.8641 Seconds
(without alignment)
1887.186 Million cell updates/sec

Title: US-09-894-912A-34
Perfect score: 1510
Sequence: 1 MHLRLISLWFLILNFMRYIG.....QQKKRKVDKQKSVSVTVH 272

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	192	12.7	1299	2 T43251	furin (EC 3.4.21.7)
2	186.5	12.4	1548	2 S34583	serine proteinase
3	184	12.2	962	2 JC5571	subtilisin-like pr
4	184	12.2	975	2 JC5570	subtilisin-like pr
5	176.5	11.7	1680	2 A43434	furin (EC 3.4.21.7)
6	173.5	11.5	969	1 A39490	subtilisin-like pr
7	168	11.1	915	2 B48225	probable proprotel
8	167.5	11.1	932	2 F52527	PACEA - mouse (fr
9	167	11.1	915	1 A48225	subtilisin-like pr
10	162	10.7	899	2 G02428	subtilisin-like pr
11	162	10.7	915	2 G02428	subtilisin-like pr
12	161	10.7	440	2 T24232	hypothetical prote
13	158.5	10.5	379	2 A59180	Wnt inhibitory fac
14	156.5	10.4	937	2 F53282	gene PACE4 protein
15	156	10.3	803	2 A47723	F-spondin precuro
16	153.5	10.2	942	2 D87803	protein bli-4D [im
17	150.5	10.0	1203	2 A49175	insulin-like growt
18	149	9.9	1371	2 A33837	insulin-like growt
19	148.5	9.8	2524	2 A35844	Xotch protein - Af
20	147.5	9.8	2471	2 A49128	cell-rate determin
21	145	9.6	570	2 T37314	probable kexin (EC
22	145	9.6	1620	2 T27283	hypothetical prote
23	143.5	9.5	807	2 A38152	F-spondin - rat
24	142	9.4	378	2 B59180	Wnt inhibitory fac
25	142	9.4	677	2 C42125	trophozoite cystei
26	139	9.2	1367	1 IGHUR1	insulin-like growt
27	139	9.2	2555	2 A40043	notch protein homo
28	138.5	9.2	1369	2 S70713	protein-tyrosine k
29	138	9.1	2531	2 S18108	notch protein homo

30	138	9.1	2531	2 A46019	notch-1 protein -
31	137	9.1	2437	2 S42612	transmembrane prot
32	136	9.0	540	2 B47417	insulin receptor-r
33	135.5	9.0	1382	1 INHUR	insulin receptor p
34	134.5	8.9	1111	2 T26972	hypothetical prote
35	134.5	8.9	1372	2 A34157	insulin receptor p
36	134.5	8.9	1383	2 A36080	insulin receptor p
37	129.5	8.6	2101	2 S57245	insulin receptor (
38	129.5	8.6	2148	1 A56081	insulin receptor -
39	128	8.5	327	2 A46484	apoptosis-mediati
40	128	8.5	1268	2 B36502	insulin receptor-r
41	127	8.4	861	2 A48825	Notch homolog Motc
42	126.5	8.4	3871	2 T22812	hypothetical prote
43	125.5	8.3	667	2 A48579	trophozoite surfac
44	125	8.3	837	2 S43656	furin (EC 3.4.21.7
45	124	8.2	1274	2 T42017	cysteine rich prot

ALIGNMENTS

RESULT 1

T43251
furin (EC 3.4.21.75) - fall armyworm
N:Alternate names: paired basic amino acid cleaving enzyme; proprotein convertase; serin
C:Species: Spodoptera frugiperda (fall armyworm)
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
C:Accession: T43251
R:Cieplik, M.; Klenk, H.
A:Description: Cloning and functional characterization of FURIN from Spodoptera frugiper
A:Reference number: 222368
A:Accession: T43251
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1299 <CIE>
A:Cross-references: EMBL:Z68888; NID:G1167859; PID:e219690; PIDN:CAA93116.1
A:Experimental source: clone Sfurin 6; ovary
C:Function:
A:Description: responsible for the endoproteolytic processing of proproteins with specif
C:Keywords: hydrolase; serine proteinase

Query Match	12.7%	Score 192;	DB 2;	Length 1299;
Best Local Similarity	27.8%	Pred. No. 2.7e-05;		
Matches	63;	Conservative	26;	Mismatches 78; Indels 60; Gaps 12;
Qy	36	NVSGCGCGCATCSD-YNGCLSCKPRUFFALERIGMKQIGVCLSCSPSGYGYTRYPDINK 94		
Db	795	SVCRPCAACATCERADGCTCERHL-----VLHDGTCMASCPSPSHYET---EDIM 843		
Qy	95	CTKCEADCTCF--NKNFCTKCKSGFYHLGKCLDNCPEGLEANNHTMECVSIHVCEVSE 152		
Db	844	CARKCHSDTCQSGPGETQCVTCHPSTYALDGRCVTSCPPAYYADKKRKEC---MRCPGV- 899		
Qy	153	WNWPSCTKKGKTCGFKGTETRVRETIQHPSAKGNLCPP---TNETRKC-TVORRKKCKQ 208		
Db	900	---CSTT-----SAPCLSCPEKWEINKKCKMPVSGDKCSA 933		
Qy	209	GERKKGRERKPKNK-----CESKE---AIPDSKSLSSKEIPE 246		
Db	934	GEPAV---DQKCKRCPACDSCYGENEGHCLTCFNPMLQDYKCVPE 977		

RESULT 2

S34583
serine proteinase (EC 3.4.21.-) PC6B - mouse
C:Species: Mus musculus (house mouse)
C>Date: 02-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 05-Nov-1999
C:Accession: S34583
R:Nakagawa, T.; Murakami, K.; Nakayama, K.
A:Title: Identification of an isoform with an extremely large Cys-rich region of PC6, a l
A:Reference number: S34583; MUID:93327934; PMID:8335106

A;Accession: S34583
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1548 <NA>
A;Cross-references: GB:D17583; NID:G407344; PIDN:BAA04507.1; PID:G405033; PID:G440374
C;Keywords: hydrolase; serine proteinase

Query Match 12.4%; Score 186.5; DB 2; Length 1548;
Best Local Similarity 25.0%; Pred. No. 6.9e-05;
Matches 66; Conservative 31; Mismatches 90; Indels 77; Gaps 15;

QY 38 SQCGCGCATCSYNG--CLSCPRLPFALLERGMKQIGVCLSSCPSSGYGTRYPI--N 93
DB 880 NQCHSSCKTC--NGSLCASCPTGMVLMQ-----ACVSPQGG-----TWPSVTSG 924
QY 94 KCTKACADCTCNKAPFTCKS-----GFTLHGKLDNCPGLEANNHTMB--CVSIVHC 148
DB 925 SCEKSEDCVSCGADLCQCLSQPONTLLHREGCVHSCPEGYAKDGVCEHCSS--PC 982
QY 149 EVSEMPWSPCTKKG-----KTC-----GFKRGTRVREIIGHPSAKGNL-- 189
DB 983 KTCGNATNSCSEGFVLDHGVCMKCPKHAHVEGVCKHCPERCODCIHEKTKCEMP 1042
QY 190 -----CPPT--NETRKTCTVQRKCKQKGERGKGRKRKRKPKNKGSKAIPDS 235
DB 1043 DFLYNDMCHRSCKSFYPMRCQVPCNK-----LECNPKREDDCKVCADTS 1091
QY 236 KSLESS---KEIPE-----QRENQK 252
DB 1092 KALHNGCLDLCEPGETYKEBDE 1115

RESULT 3
JC5571
subtilisin-like proprotein convertase (EC 3.4.21.-) PACE4 precursor, splice form E-II -
C;Species: Homo sapiens (man)
C;Date: 23-Sep-1997 #sequence_revision 23-Sep-1997 #text_change 20-Jun-2000
C;Accession: JC5571
R;Mori, K.; Kii, S.; Tsuji, A.; Nagahama, M.; Imanaki, A.; Hayashi, K.; Akamatsu, T.; Na
J. Biochem. 121, 941-948, 1997
A;Title: A novel human PACE4 isoform, PACE4E is an active processing protease containing
A;Reference number: JC5570; MUID:97335942; PMID:9192737
A;Accession: JC5571
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-962 <MOR>
A;Cross-references: DDBJ:D87994; NID:G2330550; PIDN:BAA21792.1; PID:G2330551
A;Experimental source: brain cerebellum
C;Comment: This enzyme is a processing protease and responsible for processing of various
ch it is retained intracellularly.
C;Genetics:
A;Gene: GDB:PACE4
A;Cross-references: GDB:131390; OMIM:167405
A;Map position: 15q26-15q26
C;Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
C;Keywords: glycoprotein; hydrolase; serine proteinase
F;1-62/Domain: signal sequence #status predicted <SIG>
F;63-149/Domain: propeptide #status predicted <PRO>
F;196-434/Domain: subtilisin homology <SRT>
F;938-954/Domain: hydrophobic cluster #status predicted <HCL>
F;205,246,347,420/Active site: Asp, His, Asn, Ser #status predicted
F;259/Binding site: carboxydrate (Asn) (covalent) #status predicted

Query Match 12.2%; Score 184; DB 2; Length 962;
Best Local Similarity 26.6%; Pred. No. 7.1e-05;
Matches 62; Conservative 24; Mismatches 87; Indels 60; Gaps 14;

QY 18 YIGSQNASRGRQRMRHNPVNSQCGCATCS--YNGCLSCPRLPFALLERGMKQIGV 75
DB 724 YFGDTAARRCR-----CHKGCTCSSRAATQCLSCR-RGFY-----HHQEMNT 766
QY 76 CLSSCPSSGYGTRYPIPKCKKADCTCFNK-NFTCKSGFYHLGKLDNCPGLE 134
DB 776 CLSSCPSSGYGTRYPIPKCKKADCTCFNK-NFTCKSGFYHLGKLDNCPGLE 134

767 CVTLCPAGFYADE--SQKNCLKHPSCCKVDPEKCTCKVCKBPSLARGSCIPDCEPGTY 824
QY 135 ANNHTMEC-----VSIVHCVSV--EMNPWSPCTKCKGCGFRGTETRVRE 178
DB 825 FDSLELRGCEHCHTCTGTCVGPGRBECIHCAKNFPHDM-----KCPACGSGFYPE----- 875
QY 179 IIGHPSAKGNLC-----PPTNETRKTCTVQRKCKQKGERGKGRKRKRKPKNKG 227
DB 876 --EMPGLPKVCRRYGPGE--RQATVS-----SKGVPG--GQSLSASSPGAGE 919

RESULT 4
JC5570
subtilisin-like proprotein convertase (EC 3.4.21.-) PACE4 precursor, splice form E-I -
C;Species: Homo sapiens (man)
C;Date: 23-Sep-1997 #sequence_revision 23-Sep-1997 #text_change 20-Jun-2000
C;Accession: JC5570
R;Mori, K.; Kii, S.; Tsuji, A.; Nagahama, M.; Imanaki, A.; Hayashi, K.; Akamatsu, T.; Na
J. Biochem. 121, 941-948, 1997
A;Title: A novel human PACE4 isoform, PACE4E is an active processing protease containing
A;Reference number: JC5570; MUID:97335942; PMID:9192737
A;Accession: JC5570
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-975 <MOR>
A;Cross-references: DDBJ:D87993; NID:G2330548; PIDN:BAA21791.1; PID:G2330549
A;Experimental source: brain cerebellum
C;Comment: This enzyme is a processing protease and responsible for processing of various
ch it is retained intracellularly.
C;Genetics:
A;Gene: GDB:PACE4
A;Cross-references: GDB:131390; OMIM:167405
A;Map position: 15q26-15q26
C;Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
C;Keywords: alternative splicing; glycoprotein; hydrolase; serine proteinase
F;1-62/Domain: signal sequence #status predicted <SIG>
F;63-149/Domain: propeptide #status predicted <PRO>
F;196-434/Domain: subtilisin homology <SRT>
F;952-968/Domain: hydrophobic cluster #status predicted <HCL>
F;205,246,347,420/Active site: Asp, His, Asn, Ser #status predicted
F;259/Binding site: carboxydrate (Asn) (covalent) #status predicted

Query Match 12.2%; Score 184; DB 2; Length 975;
Best Local Similarity 26.6%; Pred. No. 7.2e-05;
Matches 62; Conservative 24; Mismatches 87; Indels 60; Gaps 14;

QY 18 YIGSQNASRGRQRMRHNPVNSQCGCATCS--YNGCLSCPRLPFALLERGMKQIGV 75
DB 737 YFGDTAARRCR-----CHKGCTCSSRAATQCLSCR-RGFY-----HHQEMNT 779
QY 76 CLSSCPSSGYGTRYPIPKCKKADCTCFNK-NFTCKSGFYHLGKLDNCPGLE 134
DB 780 CVTLCPAGFYADE--SQKNCLKHPSCCKVDPEKCTCKVCKBPSLARGSCIPDCEPGTY 837
QY 135 ANNHTMEC-----VSIVHCVSV--EMNPWSPCTKCKGCGFRGTETRVRE 178
DB 838 FDSLELRGCEHCHTCTGTCVGPGRBECIHCAKNFPHDM-----KCPACGSGFYPE----- 888
QY 179 IIGHPSAKGNLC-----PPTNETRKTCTVQRKCKQKGERGKGRKRKRKPKNKG 227
DB 889 --EMPGLPKVCRRYGPGE--RQATVS-----SKGVPG--GQSLSASSPGAGE 932

RESULT 5
A43434
furin (EC 3.4.21.75) 2 - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 31-Mar-2000
C;Accession: A43434
R;Roebroek, A.J.; Creemers, J.W.; Pauli, I.G.; Kurzik-Dumke, U.; Rentrop, M.; Gateff, E.
J. Biol. Chem. 267, 17208-17215, 1992
A;Title: Cloning and functional expression of Dfurin2, a subtilisin-like proprotein proc
A;Reference number: A43434; MUID:92381036; PMID:1512259

Db 838 FDSELIIRGECCH-----TCG 853

RESULT 7

B48225

probable proprotein convertase (EC 3.4.21.-) 5 precursor - rat

N;Alternate names: PCS precursor

C;Species: Rattus norvegicus (Norway rat)

C;Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 23-Feb-1997

C;Accession: B48225

A;Lusson, J.; Vieau, D.; Hamelin, J.; Day, R.; Chretien, M.; Seidah, N.G.

Proc. Natl. Acad. Sci. U.S.A. 90, 6691-6695, 1993

A;Title: cDNA structure of the mouse and rat subtilisin/kexin-like PC5: a candidate propro-

A;Reference number: A48225; MUID:93342056; PMID:8341687

A;Accession: B48225

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-915 <LUS>

A;Cross-references: GB:L14933

C;Superfamily: subtilisin-like proteinase PACE4; subtilisin homology

F;1-34/Domain: signal sequence #status predicted <SIG>

F;13-915/Product: probable proprotein convertase 5 #status predicted <PRO>

F;117-915/Product: probable proprotein convertase 5 #status experimental <MAT>

F;164-402/Domain: subtilisin homology <SBT>

F;173,214,388/Active site: Asp, His, Ser #status predicted

Query Match 11.1%; Score 168; DB 2; Length 915;

Best Local Similarity 26.3%; Pred. No. 0.00076;

Matches 49; Conservative 25; Mismatches 54; Indels 58; Gaps 11;

Qy 35 PNVSQ-GCQG-----GCATCSDYNGCLSCPKPLPFALLERIGMKQIGVCLSSCPGSGYGYTRY 89

Db 540 PECSVEGCGDPDCHTCLHYHKLKNTR-----ICVSSCPGHP---H 682

Qy 90 PDINKTKKADCDCTCFKNKP--CPYCKSGFYHL--LGKCLDNCPGELNHNTHMBCVSI 145

Db 683 ADKKRCRCAPNCSCFSGHADQCLCKYGYFLNBTSSCVAQCPGSGYDIKKNIC--- 739

Qy 146 VHCVEVSENPWSPTCKKCTKC-GPKRGVETRVREIIQHSAGKNCILCPPTWETKCTVORK 204

Db 740 -----GKSENCCKTCTGFHNCTE-----CRGGL---SLQGSRCV--- 771

Qy 205 KCKQKE 210

Db 772 TCEDGQ 777

RESULT 8

I52527

PACE4A - mouse (fragment)

C;Species: Mus musculus (house mouse)

C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 16-Jul-1999

C;Accession: I52527

R;Hosaka, M.; Murakami, K.; Nakayama, K.

Biochem. Res. 15, 383-390, 1994

A;Title: PACE4A is a ubiquitous endoprotease that has similar but not identical substrates

A;Reference number: I52527

A;Accession: I52527

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-932 <RES>

A;Cross-references: GB:D50060; NID:9769700; PIDN:BAA08777.1; PID:9769701

C;Superfamily: subtilisin-like proteinase PACE4; subtilisin homology

F;172-410/Domain: subtilisin homology <SBT>

Query Match 11.1%; Score 167.5; DB 2; Length 932;

Best Local Similarity 24.1%; Pred. No. 0.0083;

Matches 52; Conservative 25; Mismatches 78; Indels 61; Gaps 11;

Qy 18 YIGSQNSRGRQRQRMMHNVSGCGGATCSQDYN--GCLSCPKPLPFALLERIGMKQIGV 75

Db 679 GHYHADK-KRCKKAPN- - - - - CSECFSGHGQCMSCYGYPL- - - - - NBTNSCV 723
Qy 78 SSCPSG-YGYTRYPDINKTKCKADCDTCFKNKFTCKCKSGFYHLGKCLDNCPEGLEAN 136
Db 724 THCPDGSYQDTKK- - - - - NLCKRCSEKNTCTEFHNCTECRDLGLSGSRCSVSCBGRYPN 780
Qy 137 NHTME- - - - - CVS- - - - - IVHCEVSEW- - - - - NPWSPCTKK 162
Db 781 GQDCQPCPHRFATCAGAGADGCTGEGYFMBDGRVQSCSISYFYDHSSENGYKSCCKK 840
Qy 163 GKTC- - - - - GPKKGTB- - - - - TRVREIIQHPSAKGNLCPTNETRKTIV 201
Db 841 DISCLTNGPFGKNTCPSGYLLDLGCMGALCKDATERSWAGGFMVLVKNNLC- - - 898
Qy 202 QRKCKQK 208
Db 899 QRKVLQ 905

RESULT 12

T24232
hypothetical protein R17.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T24232

R:Barlow, K.

submitted to the EMBL Data Library, March 1997

A:Reference number: Z19860

A:Accession: T24232

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-440 <WIL>

A:Cross-references: EMBL:Z92809; PIDN:CAB07269.1; GSPDB:GND00021; CESP:R17.3

A:Experimental source: clone R17

C:Genetics:

A:Gene: CESP:R17.3

A:Map position: 3

A:Introns: 36/3; 89/3; 179/2; 241/1; 306/1; 346/2

C:Superfamily: Caenorhabditis elegans hypothetical protein R17.3

Query Match 10.7%; Score 161; DB 2; Length 440;
Best Local Similarity 23.7%; Pred. No. 0.0013;
Matches 53; Conservative 30; Mismatches 87; Indels 54; Gaps 8;

Qy 58 KPLRFPALERIGMKQIGVCLSSCPSGYGYTRYPDINKTKCKADCDTCFKNKFTCKCKSG 117
Db 172 KPHLLIRYSLSKFMPPLKVTSTPLLYENRVQPNLLYLESSISECYCDEHC- - - - - 225
Qy 118 FYHLGKCLDNCPEGLEANNITMBCSVIVHCEVSEWNPSPCTKKGKGTGFRGTETVR 177
Db 226 - - - - - VTLDGDC- - - - - CSDYTFVCP- - - - - RDCVLTMDSWTQCTADNGTGI- - - - - GTOKRLR 271
Qy 178 EIIQHPSAKGNLCPTNETRKTQVKKCKGKGRERKPKNGESKBAIPDSK 237
Db 272 HVIQHAERGGACPLKEMRTCFVE- - - - - CR- - - - - PKGSLDD- - - - - 306
Qy 238 LESSKEIPEORENQOQKRRK- - - - - VQDKQKSVSVTVH 272
Db 307 ITTVALLIDYHNKTRSKIRENNIYMDLPNVAERMKATYYCVH 350

RESULT 13

A59180

Wnt inhibitory factor-1 - human

C:Species: Homo sapiens (man)

C:Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 18-Feb-2000

C:Accession: A59180

R:Heieh, J.C.; Kodjabachian, L.; Rebert, M.L.; Rattner, A.; Smallwood, P.M.; Samos, C.H.

Nature 398, 431-436, 1999

A:Title: A new secreted protein that binds to Wnt proteins and inhibits their activities

A:Reference number: A59180; MUID:99215557; PMID:10201374

A:Accession: A59180

A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-379 <HSI>

A:Cross-references: GB:AF122922; NID:94585369; PIDN:AAD25402.1; PID:94585370

Query Match 10.5%; Score 158.5; DB 2; Length 379;
Best Local Similarity 24.9%; Pred. No. 0.0017;
Matches 60; Conservative 26; Mismatches 72; Indels 83; Gaps 16;

Qy 41 COGGCAT- - - - - CSDNYNC- - - - - LSCKPLFPALERIGMKQIGVCLSS- - - - - C 80
Db 182 CPGCGRNGGFCNERRICEPCPDGFHGHPCKEALCTPRCMG- - - - - GLCVTPGFCIC 232
Qy 81 PSYGYGYTRYPDINKTKC- - - - - KADCD- - - - - TCFKNKFTCKCKSGFYHLGKCLDNCPEGLEANNH 138
Db 233 PPGFYG- - - - - VNCDKANCSFTTCFNGGTC- - - - - FY- - - - - PGKCI- - - - - CPPGLEGE- - - 271
Qy 139 TMBCVSIVHCEVSEWNPSPCTKKGKGTGFRGTETVRREIIQHPSAKGNLCPL- - - - - 192
Db 272 - - - - - QCEISKCP- - - - - QPCNNGGKCI- - - - - KSKCKSKGYQGDLCCKPVCPEGCGAHG 318
Qy 193 - - - - - TNETRKCTVQKCKGKGRERK- - - - - KPNKGSKEAIPDSKLSLESSKEIP 245
Db 319 TCHEPNKC- - - - - QCEGWHGRHCNRYEASLIHALRPAQALRQHTPSLKAERRDPP 373
Qy 246 E 246
Db 374 E 374

RESULT 14

153282

gene PACE4 protein - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 22-Jun-1999

C:Accession: 153282

R:Johnson, R.C.; Darlington, D.N.; Hand, T.A.; Bloomquist, B.T.; Maine, R.E.

Endocrinology 135, 1178-1185, 1994

A:Title: PACE4: a subtilisin-like endoprotease prevalent in the anterior pituitary and r

A:Reference number: 153282; MUID:94349873; PMID:8070361

A:Accession: 153282

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-937 <RES>

A:Cross-references: GB:L31894; NID:9496221; PIDN:AAA61987.1; PID:9496222

C:Genetics:

A:Gene: PACE4

C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology

E:177-415/Domain: subtilisin homology <SBT>

Query Match 10.4%; Score 156.5; DB 2; Length 937;
Best Local Similarity 23.6%; Pred. No. 0.0044;
Matches 51; Conservative 24; Mismatches 80; Indels 61; Gaps 11;

Qy 18 YIGSQNSRGRQRMRHNPVSGCGGCATCDYN- - - - - GCLSKPLFPALERIGMKQIGV 75
Db 705 YFGTAARRCR- - - - - CHKGETCTGRSPTOCLSCR-RGPT- - - - - HQETWT 747
Qy 76 CLSSCPSGYGYTRYPDINKTKCKADCDTCFKNKFTCKCKSGFYHLGKCLDNCPEGLE 134
Db 748 CVTLCPAGLYADESQR- - - - - CLRHPSQCKVCDEPEKSTVCKEGLARGSCIPDCBPGTY 805
Qy 135 ANNTMEC- - - - - VSIHVCEVS- - - - - EWNPSPTKKGKTCQKRGTETR 175
Db 806 FDSBLIRGECHTCTCTCVGPSREBCHCAKSPHQDKVCVPACGE- - - - - GF- - - - - 853
Qy 176 VREIIQHPSAKGNLCPTNETRKTQVKKCKGKER 211
Db 854 - - - - - YPEMPGL- - - - - PHKVCRRCDNCLSCGSSR 881

RESULT 15

A47723

Search completed: June 29, 2004, 17:03:06
Job time : 13.8641 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 16:54:24 ; Search time 8.25243 Seconds
(without alignments)
1716.232 Million cell updates/sec

Title: US-09-894-912A-34
Perfect score: 1510
Sequence: 1 MHLRLISWLFILNFMVEYG.....QQKRRKVQDKKSVSSTVTH 272

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	186.5	12.4	1877	1	PCK5 MOUSE Q04592 mus musculus
2	184.5	12.2	1696	1	PCK5 BRACL Q9nj15 branchiostoma
3	176.5	11.7	1679	1	PUR2 DROME P30432 drosophila
4	173.5	11.5	969	1	PAC4 HUMAN P29122 homo sapien
5	168	11.1	1877	1	PCK5 RAT P41413 rattus norv
6	162	10.7	913	1	PCK5 HUMAN Q92824 homo sapien
7	158.5	10.5	379	1	WTF1 HUMAN Q9y5w5 homo sapien
8	157.5	10.4	374	1	WTF1 XENLA Q9w6f8 xenopus lae
9	156.5	10.4	937	1	PAC4 RAT Q63415 rattus norv
10	156	10.3	803	1	PSPO XENLA P35447 xenopus lae
11	150.5	10.0	2470	1	NTC2 MOUSE Q35516 mus musculus
12	149	9.9	1370	1	IGIR RAT P24062 rattus norv
13	148.5	9.8	2524	1	NOTC XENLA P21783 xenopus lae
14	147.5	9.8	2471	1	NTC2 RAT Q9q30 rattus norv
15	145	9.6	943	1	BLI4 CARBL P51559 caenorhabdi
16	143.5	9.5	379	1	WTF1 MOUSE Q9wual mus musculus
17	143.5	9.5	807	1	PSFO RAT P35446 rattus norv
18	142	9.4	378	1	WTF1 BRARE Q9w6f9 brachydanio
19	142	9.4	1373	1	IGIR MOUSE P08069 homo sapien
20	139	9.2	1367	1	IGIR HUMAN P08069 homo sapien
21	139	9.2	2556	1	NTC1 HUMAN P46531 homo sapien
22	138	9.1	2531	1	NTC1 MOUSE Q01705 mus musculus
23	138	9.1	2531	1	NTC1 RAT Q07008 rattus norv
24	137	9.1	2437	1	NTC1 BRARE P46530 brachydanio
25	135.5	9.0	1382	1	INSR HUMAN P06213 homo sapien
26	134.5	8.9	1372	1	INSR MOUSE P15208 mus musculus
27	134.5	8.9	1383	1	INSR RAT P15127 rattus norv
28	133.5	8.8	2482	1	VWF FIG Q28833 sus scrofa
29	130.5	8.6	1300	1	IRR MOUSE Q9wcl4 mus musculus
30	130	8.6	2471	1	NTC2 HUMAN Q04721 homo sapien
31	129.5	8.6	2146	1	INSR DROME P09208 drosophila
32	128	8.5	327	1	TNR6 MOUSE P25446 mus musculus
33	128	8.5	1297	1	IRR_HUMAN P14616 homo sapien

34 126 8.3 687 1 VS41 GIALA P92127 giardia lam
35 125.5 8.3 667 1 TS11 GIALA Q03185 giardia lam
36 123 8.1 461 1 TRIA_PIG P50555 sus scrofa
37 123 8.1 559 1 TRAP_PLAFA P16893 plasmodium
38 123 8.1 581 1 IRR_RAT Q64716 rattus norv
39 119.5 7.9 1429 1 LI12 CARBL P14585 caenorhabdi
40 119 7.9 1300 1 IRR_CAVPO P14617 cavia porce
41 119 7.9 2703 1 NOTC DROME P07207 drosophila
42 118.5 7.8 1477 1 HTK7_HYDAT Q25197 hydra atten
43 118 7.8 867 1 SSPO_BOVIN P98167 bos taurus
44 117.5 7.8 833 1 DL_DROME P10041 drosophila
45 117.5 7.8 1367 1 LT23_CARBL P24348 caenorhabdi

ALIGNMENTS

RESULT 1
PCK5 MOUSE
ID_PCK5_MOUSE STANDARD; PRT; 1877 AA.
AC Q04592; Q62040;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Proprotein convertase subtilisin/kexin type 5 precursor (BC 3.4.21.-)
DE (Proprotein convertase PCS) (Subtilisin/kexin-like protease PCS) (PC6)
DE (Subtilisin-like proprotein convertase 6) (SPC6).
GN PCK5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE OF 330-1877 FROM N.A. (ISOFORM PCSB).
RC STRAIN=ICR; TISSUE=Intestine;
RX MEDLINE=93327934; PubMed=8335106;
RA Nakagawa T., Murakami K., Nakayama K.;
RT "Identification of an isoform with an extremely large Cys-rich region of PCS, a Kex2-like processing endoprotease.";
RT FEBS Lett. 327:165-171(1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM PCSA).
RC TISSUE=Brain, and Intestine;
RX MEDLINE=93324489; PubMed=8468318;
RA Nakagawa T., Hosaka M., Torii S., Watanabe T., Murakami K., Nakayama K.;
RT "Identification and functional expression of a new member of the mammalian Kex2-like processing endoprotease family: its striking structural similarity to PACR4.";
RT J. Biochem. 113:132-135(1993).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM PCSA).
RC TISSUE=Adrenal cortex;
RX MEDLINE=93342056; PubMed=8341687;
RA Lusson J., Vieau D., Hamelin J., Day R., Chretien M., Seidah N.G.;
RT "cDNA structure of the mouse and rat subtilisin/kexin-like PCS: a candidate proprotein convertase expressed in endocrine and nonendocrine cells.";
RT Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).
RN [4]
RP PARTIAL SEQUENCE, AND SUBCELLULAR LOCATION.
RC MEDLINE=97103178; PubMed=8947550;
RA De Bie I., Marcinkiewicz M., Malide D., Lazure C., Nakayama K., Bendayan M., Seidah N.G.;
RT "The isoforms of proprotein convertase PCS are sorted to different subcellular compartments.";
RT J. Cell Biol. 135:1261-1275(1996).
RN [5]
RP DEVELOPMENTAL EXPRESSION.
RC MEDLINE=96293359; PubMed=8698813;
RA Constam D.B., Calton M., Robertson B.J.;
RT "SPC4, SPC6, and the novel protease SPC7 are coexpressed with bone morphogenetic proteins at distinct sites during embryogenesis.";

Db 1421 KALHNGCLDCPEGTGKEENDE 1444

RESULT 2

PK5 BRACL STANDARD; PRT; 1696 AA.

ID PK5 BRACL STANDARD; PRT; 1696 AA.

AC Q9NJ15; Q9NJ14; Q9NJ16;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Proprietary convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)

DE (Proprietary convertase PCS-like) (aPC6).

GN PC6.

OS Brachyostoma californiensis (California lancelet) (Amphioxus).

OC Bukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;

OC Branchiostoma.

OX NCBI_TaxID=7738;

RP [1]

SEQUENCE FROM N.A. (ISOFORMS A; B AND C).

MDLINE=20175281; PubMed=10708868;

RA Oliva A.A. Jr., Chan S.J., Steiner D.F.;

RT "Evolution of the prohormone convertases: identification of a

RT homologue of PCS in the protochordate amphioxus";

RL Homolog. Biophys. Acta 1477:338-348(2000).

CC -1- FUNCTION: likely to represent a widespread endoprotease activity

CC within the constitutive and regulated secretory pathway. Capable

CC of cleavage at the RX(K/R)R consensus motif (By similarity).

CC -1- CATALYTIC ACTIVITY: Release of mature proteins from their

CC proproteins by cleavage of Arg-Xaa-Yaa-Arg/-Zaa bonds, where Xaa

CC can be any amino acid and Yaa is Arg or Lys.

CC -1- SUBCELLULAR LOCATION: ISOFORM A AND ISOFORM C ARE SECRETED.

CC ISOFORM B IS A TYPE I MEMBRANE PROTEIN.

CC -1- ALTERNATIVE PRODUCTS:

CC Event-Alternative splicing; Named isoforms=3;

CC Name=B;

CC IsoId=Q9NJ15-1; Sequence=Displayed;

CC Name=A;

CC IsoId=Q9NJ15-2; Sequence=VSP_005444, VSP_005445;

CC Name=C;

CC IsoId=Q9NJ15-3; Sequence=VSP_005442, VSP_005443;

CC -1- DOMAIN: The propeptide domain acts as an intramolecular chaperone

CC assisting the folding of the zymogen within the endoplasmic

CC reticulum.

CC -1- SIMILARITY: Belongs to peptidase family S8.

CC -1- SIMILARITY: Contains 1 homo B/P domain.

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

EMBL; AF184615; AAF26300.1; .

DR EMBL; AF184616; AAF26301.1; .

DR EMBL; AF184617; AAF26302.1; .

DR HSP; Q99405; IMPT.

DR MEROPS; S08.0PB; .

DR InterPro; IPR006212; Furin repeat.

DR InterPro; IPR009030; Grow fac recep.

DR InterPro; IPR00209; Peptidase_S8.

DR InterPro; IPR002884; Peptidase_S8B.

DR InterPro; IPR009020; Protease_inhib.

DR Pfam; PF01483; P_protein; 1.

DR Pfam; PF00082; Peptidase_S8; 1.

DR PRINTS; PR00723; SUBTILISIN.

DR ProDom; PD000717; P_domain; 1.

DR SMART; SM00261; FU; 17.

DR PROSITE; PS00136; SUBTILASE ASP; FALSE_NEG.

DR PROSITE; PS00137; SUBTILASE SER; 1.

DR PROSITE; PS00138; SUBTILASE SER; 1.

DR

Db 1421 KALHNGCLDCPEGTGKEENDE 1444

RESULT 2

PK5 BRACL STANDARD; PRT; 1696 AA.

ID PK5 BRACL STANDARD; PRT; 1696 AA.

AC Q9NJ15; Q9NJ14; Q9NJ16;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Proprietary convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)

DE (Proprietary convertase PCS-like) (aPC6).

GN PC6.

OS Brachyostoma californiensis (California lancelet) (Amphioxus).

OC Bukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;

OC Branchiostoma.

OX NCBI_TaxID=7738;

RP [1]

SEQUENCE FROM N.A. (ISOFORMS A; B AND C).

MDLINE=20175281; PubMed=10708868;

RA Oliva A.A. Jr., Chan S.J., Steiner D.F.;

RT "Evolution of the prohormone convertases: identification of a

RT homologue of PCS in the protochordate amphioxus.";

RL Homolog. Biophys. Acta 1477:338-348(2000).

CC -1- FUNCTION: likely to represent a widespread endoprotease activity

CC within the constitutive and regulated secretory pathway. Capable

CC of cleavage at the RX(K/R)R consensus motif (By similarity).

CC -1- CATALYTIC ACTIVITY: Release of mature proteins from their

CC proproteins by cleavage of Arg-Xaa-Yaa-Arg/-Zaa bonds, where Xaa

CC can be any amino acid and Yaa is Arg or Lys.

CC -1- SUBCELLULAR LOCATION: ISOFORM A AND ISOFORM C ARE SECRETED.

CC ISOFORM B IS A TYPE I MEMBRANE PROTEIN.

CC -1- ALTERNATIVE PRODUCTS:

CC Event-Alternative splicing; Named isoforms=3;

CC Name=B;

CC IsoId=Q9NJ15-1; Sequence=Displayed;

CC Name=A;

CC IsoId=Q9NJ15-2; Sequence=VSP_005444, VSP_005445;

CC Name=C;

CC IsoId=Q9NJ15-3; Sequence=VSP_005442, VSP_005443;

CC -1- DOMAIN: The propeptide domain acts as an intramolecular chaperone

CC assisting the folding of the zymogen within the endoplasmic

CC reticulum.

CC -1- SIMILARITY: Belongs to peptidase family S8.

CC -1- SIMILARITY: Contains 1 homo B/P domain.

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

EMBL; AF184615; AAF26300.1; -

DR EMBL; AF184616; AAF26301.1; -

DR EMBL; AF184617; AAF26302.1; -

DR HSP; Q99405; IMPT.

DR MEROPS; S08.0PB; -

DR InterPro; IPR006212; Furin repeat.

DR InterPro; IPR009030; Grow fac recep.

DR InterPro; IPR00209; Peptidase_S8.

DR InterPro; IPR002884; Peptidase_S8B.

DR InterPro; IPR009020; Protease_inhib.

DR Pfam; PF01483; P_protein; 1.

DR Pfam; PF00082; Peptidase_S8; 1.

DR PRINTS; PR00723; SUBTILISIN.

DR ProDom; PD000717; P_domain; 1.

DR SMART; SM00261; FU; 17.

DR PROSITE; PS00136; SUBTILASE ASP; FALSE_NEG.

DR PROSITE; PS00137; SUBTILASE SER; 1.

DR PROSITE; PS00138; SUBTILASE SER; 1.

DR

OC Ephydroidea; Drosophilidae; Drosophila.
 RN NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R, Tuebingen, and Iso-1;
 RX MEDLINE=92381036; PubMed=1512259;
 RA Roebroeck A.J.M., Creemers J.W.M., Pauli I.G.L., Kurzik-Dumke U.,
 RA Rentrop M., Gatteff E.A.P., Leunissen J.A.M., van de Ven W.J.M.;
 RT "Cloning and functional expression of Dfurin2, a subtilisin-like
 RT protease processing enzyme of Drosophila melanogaster with multiple
 RT repeats of a cysteine motif.";
 RL J. Biol. Chem. 267:17208-17215(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Iso-1;
 RX MEDLINE=95186060; PubMed=7880443;
 RA Roebroeck A.J.M., Ayoubi T.A.Y., Creemers J.W.M., Pauli I.G.L.,
 RA van de Ven W.J.M.;
 RT "The Dfur2 gene of Drosophila melanogaster: genetic organization,
 RT expression during embryogenesis, and pro-protein processing activity
 RT of its translational product Dfurin2.";
 RL DNA Cell Biol. 14:223-234 (1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Baillew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:1215-2195(2000).
 CC -!- FUNCTION: Furin is likely to represent the ubiquitous endoprotease
 CC activity within constitutive secretory pathways and capable of
 CC cleavage at the RX(K/R)R consensus motif (By similarity).
 CC -!- CATALYTIC ACTIVITY: Release of mature proteins from their
 CC proproteins by cleavage of Arg-Xaa-Yaa-Arg|-Zaa bonds, where Xaa
 CC can be any amino acid and Yaa is Arg or Lys. Releases albumin,
 CC complement component C3 and von Willebrand factor from their
 CC respective precursors.
 CC -!- TISSUE SPECIFICITY: Transient expression in a subset of central

CC nervous system neurons during embryonic stages 12-13. Expression
 CC in developing tracheal tree from stage 13 to end of embryonic
 CC development.
 CC -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.
 CC -!- SIMILARITY: Belongs to peptidase family S8. Furin subfamily.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M94375; AAA28551.1; --
 CC EMBL; L33831; AAA69860.1; --
 CC EMBL; AB003502; AAP48598.1; --
 CC FIR; A43434; A43434.
 CC HSSP; Q99405; IMPT.
 CC -----
 CC FLYBASE: FBgn0004598; Fur2.
 CC GO; GO:0004276; P: furin activity; IDA.
 CC InterPro; IPR006212; Furin repeat.
 CC InterPro; IPR009030; Grow_fac_recep.
 CC InterPro; IPR000209; Peptidase_S8.
 CC InterPro; IPR002884; Peptidase_S8B.
 CC InterPro; IPR009020; Protease_inhib.
 CC Pfam; PF01483; P_proteasein; 2.
 CC Pfam; PF00082; Peptidase_S8; 1.
 CC PRINTS; PR00723; SUBTILISIN.
 CC ProDom; PD000717; P_domain; 1.
 CC SMART; SM00261; FU; 10.
 CC PROSITE; PS00136; SUBTILASE_ASP; 1.
 CC PROSITE; PS00137; SUBTILASE_HIS; 1.
 CC PROSITE; PS00138; SUBTILASE_SER; 1.
 CC HydroLase; Serine protease; Glycoprotein; Signal; Transmembrane;
 CC Multigene family; Zymogen; Repeat.
 CC SIGNAL 1 ? POTENTIAL.
 CC PROPEP 2 318 POTENTIAL.
 CC CHAIN 319 1679 FURIN-LIKE PROTEASE 2.
 CC ACT_SITE 417 417 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC ACT_SITE 456 456 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC ACT_SITE 637 637 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC DOMAIN 961 1443 10 X TANDEM REPEATS, CYS-RICH.
 CC REPEAT 961 1006 1.
 CC REPEAT 1007 1056 2.
 CC REPEAT 1057 1103 3.
 CC REPEAT 1104 1152 4.
 CC REPEAT 1153 1204 5.
 CC REPEAT 1205 1253 6.
 CC REPEAT 1254 1298 7.
 CC REPEAT 1299 1345 8.
 CC REPEAT 1346 1392 9.
 CC REPEAT 1393 1443 10.
 CC TRANSMEM 1512 1532 POTENTIAL.
 CC DOMAIN 1533 1679 CYTOPLASMIC (POTENTIAL).
 CC CARBOHYD 3 3 N-LINKED (GLCNAC...) (POTENTIAL).
 CC CARBOHYD 109 109 N-LINKED (GLCNAC...) (POTENTIAL).
 CC CARBOHYD 130 130 N-LINKED (GLCNAC...) (POTENTIAL).
 CC CARBOHYD 205 205 N-LINKED (GLCNAC...) (POTENTIAL).
 CC CARBOHYD 442 442 N-LINKED (GLCNAC...) (POTENTIAL).
 CC CARBOHYD 480 480 N-LINKED (GLCNAC...) (POTENTIAL).
 CC CARBOHYD 927 927 N-LINKED (GLCNAC...) (POTENTIAL).
 CC CARBOHYD 1060 1060 N-LINKED (GLCNAC...) (POTENTIAL).
 CC CARBOHYD 1181 1181 N-LINKED (GLCNAC...) (POTENTIAL).
 CC CARBOHYD 1274 1274 N-LINKED (GLCNAC...) (POTENTIAL).
 CC CARBOHYD 1277 1277 N-LINKED (GLCNAC...) (POTENTIAL).
 CC CARBOHYD 1439 1439 N-LINKED (GLCNAC...) (POTENTIAL).
 CC CONFLICT 152 153 MISSING (IN REF. 1).
 CC CONFLICT 177 177 V -> F (IN REF. 1).
 CC CONFLICT 213 213 V -> VDQL (IN REF. 1).
 CC SEQUENCE 1679 AA; 183369 MW; 3F9E749F0B021CF6 CRC64;

```

Query Match      11.78; Score 176.5; DB 1; Length 1679;
Best Local Similarity 28.08; Pred. No. 2.3e-05;
Matches 60; Conservative 24; Mismatches 77; Indels 53; Gaps 12;

Qy 2 HLRILSWLFIILNFWBYIGSQNASRGRQRMRHPNVSGCGGCATCSDY-NGCLSCCKPR 60
Db 1037 HLAVID-LAVCLQCPDGPFPNS-----RNRTCVF-----CEPNCASQDHPFYCTSCDHH 1086

Qy 61 LPFALEIRIGHKQIGVCLSSCPGSGYGYTRYDINKTKTKKADCTCP--NNFCTKCKSGF 118
Db 1087 LVNHEHK-----CYSACPLDTYET---EDNRCAPFCHSTCATCGPTDQDCITCRSSR 1135

Qy 119 YLHLGKCLDNCPEGLANNHMECVSVHCVSEWNPSPCTKGGKTCGPKRGKTETRV-R 177
Db 1136 YAWQNKCLISCPDGFYADKKKLECH-----PCQEGCKTC-----TSNGVCS 1176

Qy 178 EIIQHPKAGNLCPTTNETRCKTVQRK-CQKGE 210
Db 1177 ECLQNTI-----LNKRDKCIVSGEGCSESE 1202

RESULT 4
PAC4 HUMAN STANDARD; PRT; 969 AA.
ID PAC4 HUMAN STANDARD; PRT; 969 AA.
AC P29122; Q15099; Q15100; Q9UEJ7; Q9UEJ2; Q9UEJ3; Q9UEJ8;
AC Q9UEJ9; Q9Y4G9; Q9Y4H0; Q9Y4H1;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Paired basic amino acid cleaving enzyme 4 precursor (EC 3.4.21.-)
DE (Subtilisin/kexin-like protease PACB4) (Subtilisin-like proprotein
DE convertase 4) (SPC4).
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (ISOFORMS PACB4A-I AND PACB4B).
RC TISSUE=Hepatoma, and Kidney;
RX MEDLINE=92075167; PubMed=1741956;
Kiefer M.C., Tucker J.E., Joh R., Landsberg K.E., Saltman D.,
Barr P.J.;
RA Identification of a second human subtilisin-like protease gene in
RT the tes/fps region of chromosome 15.;
RL DNA Cell Biol. 10:757-769(1991).
[2]
SEQUENCE FROM N.A. (ISOFORMS PACB4C AND PACB4D).
RC TISSUE=Placenta;
RX MEDLINE=94235049; PubMed=8179631;
Teuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,
Matsuda Y.;
RA Identification of novel cDNAs encoding human kexin-like protease,
RT PACB4 isoforms.;
RL Biochem. Biophys. Res. Commun. 200:943-950(1994).
[3]
ERRATUM.
RX MEDLINE=95071480; PubMed=7980617;
Teuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,
Matsuda Y.;
RA Identification of novel cDNAs encoding human kexin-like protease,
RT PACB4 isoforms.;
RL Biochem. Biophys. Res. Commun. 204:1381-1382(1994).
[4]
SEQUENCE FROM N.A. (ISOFORM PACB4A-II).
RC TISSUE=Placenta;
RA Mori K., Imamaki A., Kii S., Nagamune H., Nagahama M., Tsuji A.,
Matsuda Y.;
RT Identification of a novel PACB4 isoform, PACB4E.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
[5]
SEQUENCE FROM N.A. (ISOFORMS PACB4E-I AND PACB4E-II).
RC TISSUE=Cerebellum;

```

```

RX MEDLINE=97335942; PubMed=9192737;
RA Mori K., Kii S., Tsuji A., Nagahama M., Imamaki A., Hayashi K.,
RA Akamatsu T., Nagamune H., Matsuda Y.;
RT "A novel human PACB4 isoform, PACB4E is an active processing protease
RL containing a hydrophobic cluster at the carboxy terminus.";
RL J. Biochem. 121:941-948(1997).
[6]
SEQUENCE FROM N.A. (ISOFORMS PACB4A-I; PACB4A-II; PACB4CS; PACB4D;
PACB4E-I AND PACB4E-II).
RX MEDLINE=98021085; PubMed=9378725;
RA Teuji A., Hine C., Tamai Y., Yonemoto K., Mori K., Yoshida S.,
RA Bando M., Sakai E., Mori K., Akamatsu T., Matsuda Y.;
RA "Genomic organization and alternative splicing of human PACB4 (SPC4),
RT kexin-like processing endoprotease.";
RL J. Biochem. 122:438-452(1997).
[7]
ALTERNATIVE SPLICING (ISOFORM PACB4CS).
RX MEDLINE=97064242; PubMed=8906861;
Zhang M., Benjannet S., Lazure C., Munzer S., Seidah N.G.;
RT "Functional analysis of human PACB4-A and PACB4-C isoforms:
RT identification of a new PACB4-CS isoform.";
RL FEBS Lett. 396:31-36(1996).
[8]
CHARACTERIZATION.
RX MEDLINE=99233559; PubMed=10215603;
Susic J.P., Moehring J.M., Innocencio N.M., Luchini J.M.,
RA Moehring T.J.;
RA "Endoprotease PACB4 is Ca2+-dependent and temperature-sensitive and
RT can partly rescue the phenotype of a furin-deficient cell strain.";
RL Biochem. J. 339:639-647(1999).
[9]
PROCESSING.
RX MEDLINE=98408849; PubMed=9738469;
Nagahama M., Taniguchi T., Hashimoto E., Imamaki A., Mori K.,
RA Teuji A., Matsuda Y.;
RA "Biosynthetic processing and quaternary interactions of proprotein
RT convertase SPC4 (PACB4).";
RL FEBS Lett. 434:155-159(1998).
[1]
FUNCTION: Likely to represent an endoprotease activity within the
CC constitutive secretory pathway, with unique restricted
CC distribution in both neuroendocrine and non-neuroendocrine tissues
CC and capable of cleavage at the RK(R/R)R consensus motif.
CC
CC -1- CATALYTIC ACTIVITY: Release of mature proteins from their
CC proproteins by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa
CC can be any amino acid and Yaa is Arg or Lys.
CC
CC -1- COFACTOR: Calcium (Potential).
CC
CC -1- SUBUNIT: The PACB4A-I precursor protein seems to exist in the
CC reticulum endoplasmic as both a monomer and a dimer-sized complex
CC whereas mature PACB4A-I exists only as a monomer, suggesting that
CC propeptide cleavage affects its tertiary or quaternary structure.
CC
CC -1- SUBCELLULAR LOCATION: PACB4A-I and PACB4A-II are secreted. PACB4C
CC and PACB4CS are not secreted and remain probably in zymogen form
CC in endoplasmic reticulum. PACB4E-I and PACB4E-II are retained
CC intracellularly probably through a hydrophobic cluster in their C-
CC terminus. PACB4B might be secreted.
CC
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=8;
CC Name=PACB4A-I; Synonyms=PACB4;
CC IsoId=P29122-1; Sequence=Displayed;
CC Name=PACB4A-II;
CC IsoId=P29122-2; Sequence=VSP_005436;
CC Name=PACB4B; Synonyms=PACB4.1;
CC IsoId=P29122-3; Sequence=VSP_005428, VSP_005429;
CC Note=Probably enzymatically inactive;
CC Name=PACB4C;
CC IsoId=P29122-4; Sequence=VSP_005432, VSP_005433;
CC Note=Probably enzymatically inactive;
CC Name=PACB4CS;
CC IsoId=P29122-5; Sequence=VSP_005430, VSP_005431;
CC Note=Probably enzymatically inactive;
CC Name=PACB4D;
CC IsoId=P29122-6; Sequence=VSP_005427, VSP_005434, VSP_005435;
CC Note=Probably enzymatically inactive;

```

CC CC Name=PAC34E-I; ISOID=P29122-7; Sequence=VSP_005437;
 CC CC Name=PAC34E-II; ISOID=P29122-8; Sequence=VSP_005436, VSP_005437;
 CC CC TISSUE SPECIFICITY: Each PAC34 isoform exhibits a unique
 CC CC restricted distribution. PAC34A-I is expressed in heart, brain,
 CC CC placenta, lung, skeletal muscle, kidney, pancreas, but at
 CC CC comparatively higher levels in the liver. PAC34A-II is at least
 CC CC expressed in placenta. PAC34B was only found in the embryonic
 CC CC kidney cell line from which it was isolated. PAC34C and PAC34D are
 CC CC expressed in placenta. PAC34E-I is expressed in cerebellum,
 CC CC placenta and pituitary. PAC34E-II is at least present in
 CC CC cerebellum.
 CC CC -!- DOMAIN: The propeptide domain acts as an intramolecular chaperone
 CC CC assisting the folding of the zymogen within the endoplasmic
 CC CC reticulum. Isoform PAC34D lacks the propeptide domain.
 CC CC -!- SIMILARITY: Belongs to peptidase family 58.
 CC CC -!- SIMILARITY: Contains 1 homo B/P domain.
 CC CC -!- SIMILARITY: Contains 1 PLAC domain.
 CC CC -----
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC CC the European Bioinformatics Institute. There are no restrictions on its
 CC CC use by non-profit institutions as long as its content is in no way
 CC CC modified and this statement is not removed. Usage by and for commercial
 CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC CC or send an email to license@isb-sib.ch).
 CC CC -----
 CC CC EMBL; MB0482; AAA59998.1; -!
 CC CC EMBL; AB001914; BAA21620.1; JOINED.
 CC CC EMBL; AB001898; BAA21620.1; JOINED.
 CC CC EMBL; AB001900; BAA21620.1; JOINED.
 CC CC EMBL; AB001901; BAA21620.1; JOINED.
 CC CC EMBL; AB001902; BAA21620.1; JOINED.
 CC CC EMBL; AB001903; BAA21620.1; JOINED.
 CC CC EMBL; AB001904; BAA21620.1; JOINED.
 CC CC EMBL; AB001905; BAA21620.1; JOINED.
 CC CC EMBL; AB001914; BAA21621.1; JOINED.
 CC CC EMBL; AB001898; BAA21621.1; JOINED.
 CC CC EMBL; AB001900; BAA21621.1; JOINED.
 CC CC EMBL; AB001901; BAA21621.1; JOINED.
 CC CC EMBL; AB001902; BAA21621.1; JOINED.
 CC CC EMBL; AB001903; BAA21621.1; JOINED.
 CC CC EMBL; AB001904; BAA21621.1; JOINED.
 CC CC EMBL; AB001905; BAA21621.1; JOINED.
 CC CC EMBL; AB001906; BAA21621.1; JOINED.
 CC CC EMBL; AB001907; BAA21621.1; JOINED.
 CC CC EMBL; AB001908; BAA21621.1; JOINED.
 CC CC EMBL; AB001909; BAA21621.1; JOINED.
 CC CC EMBL; AB001914; BAA21622.1; JOINED.
 CC CC EMBL; AB001901; BAA21622.1; JOINED.
 CC CC EMBL; AB001902; BAA21622.1; JOINED.
 CC CC EMBL; AB001903; BAA21622.1; JOINED.
 CC CC EMBL; AB001904; BAA21622.1; JOINED.
 CC CC EMBL; AB001905; BAA21622.1; JOINED.
 CC CC EMBL; AB001906; BAA21622.1; JOINED.
 CC CC EMBL; AB001907; BAA21622.1; JOINED.
 CC CC EMBL; AB001908; BAA21622.1; JOINED.
 CC CC EMBL; AB001914; BAA21623.1; JOINED.
 CC CC EMBL; AB001898; BAA21623.1; JOINED.
 CC CC EMBL; AB001900; BAA21623.1; JOINED.
 CC CC EMBL; AB001901; BAA21623.1; JOINED.
 CC CC EMBL; AB001902; BAA21623.1; JOINED.
 CC CC EMBL; AB001903; BAA21623.1; JOINED.
 CC CC EMBL; AB001904; BAA21623.1; JOINED.
 CC CC EMBL; AB001905; BAA21623.1; JOINED.
 CC CC EMBL; AB001906; BAA21623.1; JOINED.
 CC CC EMBL; AB001907; BAA21623.1; JOINED.
 CC CC EMBL; AB001908; BAA21623.1; JOINED.
 CC CC EMBL; AB001909; BAA21623.1; JOINED.
 CC CC EMBL; AB001914; BAA21624.1; JOINED.
 CC CC EMBL; AB001898; BAA21624.1; JOINED.

DR EMBL; AB001901; BAA21624.1; JOINED.
 DR EMBL; AB001902; BAA21624.1; JOINED.
 DR EMBL; AB001903; BAA21624.1; JOINED.
 DR EMBL; AB001904; BAA21624.1; JOINED.
 Query Match 11.5%; Score 173.5; DB 1; Length 969;
 Best Local Similarity 28.8%; Pred. No. 2.2e-05;
 Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;
 QY 18 YIGSNASRGRRQRMRHPNVSCQCGCATCSD--YNGCLSKRPLFFAALRIGMKQIGV 75
 DB 737 YFGDTAARCR-----CHKGCETSSRAATQCLSCR-RGFF-----HHQSMNT 779
 QY 76 CLSSCPGYYGTRYPDINKCTKACDCTCFNK-NECTKCKSGFYHLHGLCLDNCPGLE 134
 DB 780 CYTLCPAGFYADE--SQKNCLKHPCKCKVDPEKTCVCKRGFSLARGSCIPDEPGTY 837
 QY 135 ANNHTECVSIVHCEVSEWNPSPCTKKGKTCG 167
 DB 838 FDSLEIRCGECH-----TCG 853
 RESULT 5
 PKCS RAT
 ID PKCS RAT STANDARD; PRT; 1877 AA.
 AC P41413; Q62914;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Proprotein convertase subtilisin/kexin type 5 precursor (BC 3.4.21.-)
 DE (Proprotein convertase PCS) (Subtilisin/kexin-like protease PCS)
 DE (PC5) (PC6) (Fragment).
 GN PCSK5.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM SHORT), AND TISSUE SPECIFICITY.
 RC TISSUE=Adrenal gland;
 RX MEDLINE=93342056; PubMed=8341687;
 RA Lusson J., Vieau D., Hamelin J., Day R., Chretien M., Seidah N.G.;
 RT "CDNA structure of the mouse and rat subtilisin/kexin-like PCS: a
 RT candidate proprotein convertase expressed in endocrine and
 RT nonendocrine cells."
 RL Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).
 RN [2]
 RP SEQUENCE OF 1676-1877 FROM N.A. (ISOFORM LONG).
 RC TISSUE=Adrenal gland;
 RA De Bie I., Marcinkiewicz M., Nakayama K., Lazure C., Seidah N.G.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=20214819; PubMed=10749928;
 RA Xiang Y., Molloy S.S., Thomas L., Thomas G.;
 RT "The PC6B cytoplasmic domain contains two acidic clusters that direct
 RT sorting to distinct trans-Golgi network/endosomal compartments."
 RL Mol. Biol. Cell 11:1257-1273(2000).
 RN [4]
 RP DEVELOPMENTAL EXPRESSION.
 RX MEDLINE=97166043; PubMed=9013936;
 RA Zheng M., Seidah N.G., Pintar J.E.;
 RT "The developmental expression in the rat CNS and peripheral tissues of
 RT proteases PCS and PACE4 mRNAs: comparison with other proprotein
 RT processing enzymes."
 RL Dev. Biol. 181:268-283(1997).
 CC -!- FUNCTION: Likely to represent a widespread endoprotease activity
 CC within the constitutive and regulated secretory pathway. Capable
 CC of cleavage at the RX(K/R)R consensus motif. May be responsible
 CC for the maturation of gastrointestinal peptides. May be involved
 CC in the cellular proliferation of adrenal cortex via the activation
 CC of growth factors.
 CC -!- CATALYTIC ACTIVITY: Release of mature proteins from their

FT	DOMAIN	1825	1844	AC 1.
FT	DOMAIN	1856	1877	AC 2.
FT	SITE	116	117	CLEAVAGE (AUTO-) (BY SIMILARITY).
FT	SITE	521	523	CELL ATTACHMENT SITE (POTENTIAL).
FT	ACT_SITE	173	173	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	214	214	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	388	388	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	CARBOHYD	227	227	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	383	383	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	667	667	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	754	754	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	804	804	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	854	854	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1710	1710	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1732	1732	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARSPLIC	878	915	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX -->
FT				ATRESWAEGGFCMLVTKNNLCQRKVLQLCCKCTFQG
FT				(in isoform PCSA).
FT				/FTid=VSP_005440.
FT	VARSPLIC	916	1877	Missing (in isoform PCSA).
FT				/FTid=VSP_005441.
FT	SEQUENCE	1877 AA;	207888 MW;	890955DC60534444 CRC64;
Query Match				
Best Local Similarity				
Matches 49; Conservative 25; Mismatches 54; Indels 58; Gaps 11				
Qy	35	PNVSQ-GCOG----	GCATCSDYNGCLSKPRLPFALERIGMKQIGVCLSSCPSPGYGTRY	89
Db	640	PESEVGCDGPGPHCHTCLHYHYKLNTR-----	-----ICVSSCPGPHF---H	682
Qy	90	PDINKTKKACDCITCFKNPF--CTKCKSGPYLH--LGRCLDNCPEGLAANNHTMBCVSI	145	
Db	683	ADKKCRKCAPNCBSCFCGSHADQCLSKYGVFLNEETSSCVAQCPESGYQDIKKNIC---	739	
Qy	146	VHCEVSENPWSPTCKGKTC-GFGRGTETVRRIHQPSAKGMLCPPTNETRCKTQVRK	204	
Db	740	-----GKCSNCKTCTGFHNCTE-----CKGGL---	SLQSSRCSV---771	
Qy	205	KCKOGE 210		
Db	772	TCDEGQ 777		
RESULT 6				
PKCS_HUMAN				
ID	PKCS_HUMAN	STANDARD;	PRF;	913 AA.
AC	Q92824; Q13527;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Proteinase convertase subtilisin/kexin type 5 precursor (BC 3.4.21.-)			
DE	(Proteinase convertase PCS) (subtilisin/kexin-like protease PCS) (PC6)			
DE	(hPC6).			
GN	PKCS5 OR PC5 OR PC6.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_Taxid=9606;			
RN	[1]			
RN	SEQUENCE FROM N.A.			
RC	TISSUE=T-cell;			
RX	MEDLINE=96353890; PubMed=8755538;			
RL	Miranda L., Wolf J., Pichuanes S., Duke R., Franzusoff A.;			
RT	"Isolation of the human PC6 gene encoding the putative host protease			
RT	for HIV-1 gp160 processing in CD4+ T lymphocytes.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 93:7695-7700 (1996).			
RL	[2]			
RP	REVISIONS.			
RA	Franzusoff A., Miranda L., Wolf J., Pichuanes S., Lu Y., Duke R.;			
RL	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.			
RL	[3]			
RP	SEQUENCE OF 15-913 FROM N.A.			

RA Reudelhuber T.L.; to the EMBL/GenBank/DBJ databases.
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Likely to represent a widespread endoprotease activity
 CC within the constitutive and regulated secretory pathway. Capable
 CC of cleavage at the RX(K/R)R consensus motif.
 CC -!- CATALYTIC ACTIVITY: Release of mature proteins from their
 CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa
 CC can be any amino acid and Yaa is Arg or Lys.
 CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=1;
 CC Comment=2 isoforms may be produced;
 CC Name=1;
 CC IsoId=Q92824-1; Sequence=Displayed;
 CC TISSUE SPECIFICITY: Expressed in T-lymphocytes.
 CC -!- DOMAIN: The propeptide domain acts as an intramolecular chaperone
 CC assisting the folding of the zymogen within the endoplasmic
 CC reticulum.
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC -!- SIMILARITY: Contains 1 homo B/P domain.
 CC -!- SIMILARITY: Contains 1 PLAC domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U56387; AAC50643.2; -;
 CC EMBL: U49114; AAA91807.1; -;
 CC PIR: G02428; G02428.
 CC HSP: Q99405; IMPT.
 CC MEROPS: S08.076; -;
 CC Genew: HGNC:8747; PCSK5.
 CC MIT: 600488; -;
 CC GO: GO:0005615; C:extracellular space; TAS.
 CC GO: GO:0004252; F:serine-type endopeptidase activity; TAS.
 CC GO: GO:0007267; P:cell-cell signaling; TAS.
 CC GO: GO:0006508; P:proteolysis and peptidolysis; TAS.
 CC InterPro: IPR006212; Furin repeat.
 CC InterPro: IPR000209; Peptidase S8.
 CC InterPro: IPR002884; Peptidase_S8B.
 CC InterPro: IPR009020; Protease_inhib.
 CC Pfam: PF01483; P:protease; 1.
 CC Pfam: PF00082; Peptidase S8; 1.
 CC PRINTS: PR00723; SUBTILISIN.
 CC ProDom: PD000717; P domain; 1.
 CC SMART: SM00261; FU_5.
 CC PROSITE: PS00900; PLAC; 1.
 CC PROSITE: PS00136; SUBTILASE_ASP; 1.
 CC PROSITE: PS00137; SUBTILASE_HIS; 1.
 CC PROSITE: PS00138; SUBTILASE_SER; 1.
 CC Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
 CC Cleavage on pair of basic residues; Repeat; Alternative splicing.
 KW SIGNAL 1 32
 FT PROPEP 33 114
 FT CHAIN 115 913
 FT TYPE 5.
 FT PROPEPTIN CONVERTASE SUBTILISIN/KEXIN
 FT CATALYTIC.
 FT HOMO B.
 FT CYS-RICH MOTIF (CRM) REGION.
 FT PLAC.
 FT CLEAVAGE (AUTO-) (BY SIMILARITY).
 FT CELL ATTACHMENT SITE (POTENTIAL).
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 454
 FT DOMAIN 462 600
 FT DOMAIN 636 868
 FT DOMAIN 869 913
 FT SITE 114 115
 FT SITE 519 521
 FT ACT_SITE 171 171
 FT ACT_SITE 212 212
 FT ACT_SITE 386 386
 FT ACT_SITE 225 225
 FT CARBOHYD 381 381
 FT CARBOHYD 665 665
 FT CARBOHYD 752 752

FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 852 852 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 118 118 S -> F (IN REF. 3).
 FT CONFLICT 121 121 V -> A (IN REF. 3).
 FT CONFLICT 511 511 R -> A (IN REF. 3).
 FT CONFLICT 601 601 R -> Q (IN REF. 3).
 SQ SEQUENCE 913 AA; 101775 MW; 21389264CAD7546C CRC64;
 Query Match 10.7%; Score 162; DB 1; Length 913;
 Best Local Similarity 22.3%; Pred. No. 0.00014;
 Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;
 Name=1;
 QY 20 GSONASGRORRMHNVNVSQCGCGCATCDSDYNG--CLSCPKRLFPFALERGMQIGVCL 77
 DB 677 GHYADK-KRCRCAPN-----CESCFGSHGDCMSCKYGYFL-----NEETNSCV 721
 QY 78 SSCPSG-YGYTRYPDINKTKKADCTCPNKNPCTCKSGFYHLKGLDNCPEGLAN 136
 DB 722 THCPDGSYQDTKK---NLCKKSENCKTCTEFNCTECRDGLSLQGSRCVSCEBGRYFN 778
 QY 137 NHTME-----CVS-----IVHCEVSEW-----NPHSPCTKK 162
 DB 779 GDCQCPCHRCATCAGAGAGDCINCTEGYFMEDGRCVQSCSIYYPHSSSENGYKSCCK 838
 QY 163 GKTC-----GFKRGYE-----TVREIIQHPSAKGNLCPTNETRKTCTV 201
 DB 839 DISLCITGNGFPNCTSCPSGYLLDGMQGAICKDATESWAEFGFCLVKKNLCC-- 896
 QY 202 QRKKCKQK 208
 DB 897 QRKVLQK 903
 RESULT 7
 ID WIP1 HUMAN STANDARD; PRT; 379 AA.
 AC Q9Y5W5; Q8WVG4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DB Wnt inhibitory factor 1 precursor (WIF-1).
 GN WIF1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99215557; PubMed=10201374;
 RA Hsieh J.-C., Kodjabachian L., Rebbert M.L., Rattner A.,
 RA Smallwood P.M., Samos C.H., Nusse R., David I.B., Nathans J.;
 RT "A new secreted protein that binds to Wnt proteins and inhibits their
 RT activities.";
 RL Nature 398:431-436(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=22398257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
 RA Diatchenko L., Narusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Scapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.B.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A.C., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A.C., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

DR PROSITE; PS01186; EGF 2; 5.
 DR PROSITE; PS50026; EGF 3; 4.
 DR PROSITE; PS50814; WIF; 1.
 KW Repeat; EGF-like domain; Signal; Developmental protein;
 KW Wnt signaling pathway.
 FT SIGNAL 1 28
 FT CHAIN 29 374
 FT DOMAIN 33 171
 FT DOMAIN 172 203
 FT DOMAIN 204 235
 FT DOMAIN 236 267
 FT DOMAIN 268 299
 FT DOMAIN 300 331
 FT DISULFID 172 181
 FT DISULFID 177 187
 FT DISULFID 193 195
 FT DISULFID 204 213
 FT DISULFID 209 219
 FT DISULFID 225 227
 FT DISULFID 236 245
 FT DISULFID 241 251
 FT DISULFID 257 259
 FT DISULFID 268 277
 FT DISULFID 273 283
 FT DISULFID 289 291
 FT DISULFID 300 309
 FT DISULFID 305 315
 FT DISULFID 321 323
 FT CARBOHYD 83 83
 FT CARBOHYD 240 240
 SQ SEQUENCE 374 AA; 41071 MW; 82697380P00ACPF8 CRC64;
 Query Match 10.4%; Score 157.5; DB 1; Length 374;
 Best Local Similarity 24.3%; Pred. No. 0.00013;
 Matches 57; Conservative 22; Mismatches 85; Indels 71; Gaps 12;
 QY 41 CQGGCAT---CSDYNGC-----LSCKPRLFFALERIGMKQIGVC-----LSSC 80
 DB 177 CTGGCGNGFCNDRHVCCEPDGFGPHCEKALCMPCMG-----GLCVTPGLCIC 227
 QY 81 PSQYVTRVPDINKTKKACDCTCNKNFCTCKSGFYHLGKGLDNCPEGLRANNHTM 140
 DB 228 PPGYVINGCDKVNCTTHC-LANGTGF-----YPGKCI--CPSGYEGE----- 266
 QY 141 ECVSIVHCVSEWNPSPTCKTKTGFPRGTETVRBIIQHPSAKGNLCPPTNETRKT 200
 DB 267 -----QCETSKCQ--QPCRNKGKCSG---KNCKCKSGYQGLCKPVCPCSCGAHGTC 315
 QY 201 VQRKKCKQGERGKGRERK-----KPNKGSKEAIPDSKSLESKBIPE 246
 DB 316 IEPNKCQCKE-GWNGRYCNKYSNLNMLALPTGSRNRQHTPSPKRTEDQALPE 369
 RESULT 9
 PAC4 RAT
 ID PAC4 RAT STANDARD; PRT; 937 AA.
 AC Q63415;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Paired basic amino acid cleaving enzyme 4 precursor (EC 3.4.21.-)
 DE [Subtilisin/kexin-like protease PACE4] (Subtilisin-like propeptide
 DE convertase 4) (SPC4).
 GN PACE4.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Hypothalamus, and Pituitary;
 RX MEDLINE=94349873; PubMed=8070361;
 RA Johnson R.C., Darlington D.N., Hand T.A., Bloomquist B.T., Mains R.E.;

RT "PACE4: a subtilisin-like endoprotease prevalent in the anterior
 RT pituitary and regulated by thyroid status."; Endocrinology 135:1178-1185 (1994).
 RL CC -|- FUNCTION: Likely to represent an endoprotease activity within the
 CC constitutive secretory pathway, with unique restricted
 CC distribution in both neuroendocrine and non-neuroendocrine tissues
 CC and capable of cleavage at the RX(K/R)R consensus motif.
 CC -|- CATALYTIC ACTIVITY: Release of mature proteins from their
 CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa
 CC can be any amino acid and Yaa is Arg or Lys.
 CC -|- COFACTOR: Calcium (POTENTIAL).
 CC -|- TISSUE SPECIFICITY: High expression in the anterior pituitary and
 CC in several brain regions, the atrium, and the ventricle.
 CC -|- DOMAIN: The propeptide domain acts as an intramolecular chaperone
 CC assisting the folding of the zymogen within the endoplasmic
 CC reticulum.
 CC -|- SIMILARITY: Belongs to peptidase family S8.
 CC -|- SIMILARITY: Contains 1 homo B/P domain.
 CC -|- SIMILARITY: Contains 1 PLAC domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>)
 CC or send an email to license@ebi.ac.uk.
 CC -----
 DR EMBL; L31894; AAA61987.1; -.
 DR F1R; I53282; I53282.
 DR HSSP; Q99405; 1MPT.
 DR MEROPS; S08.075; -.
 DR InterPro; IPR006212; Furin repeat.
 DR InterPro; IPR000209; Peptidase_S8.
 DR InterPro; IPR002884; Peptidase_S8B.
 DR InterPro; IPR009020; Protease_inhib.
 DR Pfam; PF01483; P_protein; 1.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR ProDom; PD000717; P_domain; 1.
 DR SMART; SM00261; FU; 5.
 DR PROSITE; PS00900; PLAC; 1.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal; Calcium;
 KW Cleavage on pair of basic residues; Repeat.
 FT SIGNAL 1 45
 FT PROPEP 46 132
 FT CHAIN 133 937
 FT DOMAIN 133 454
 FT DOMAIN 477 615
 FT DOMAIN 680 898
 FT DOMAIN 899 937
 FT SITE 132 133
 FT SITE 534 536
 FT ACT_SITE 186 186
 FT ACT_SITE 227 227
 FT ACT_SITE 401 401
 FT CARBOHYD 240 240
 FT CARBOHYD 882 882
 FT CARBOHYD 900 900
 SQ SEQUENCE 937 AA; 104053 MW; F3865557C33705C8 CRC64;
 Query Match 10.4%; Score 156.5; DB 1; Length 937;
 Best Local Similarity 23.6%; Pred. No. 0.00036;
 Matches 51; Conservative 24; Mismatches 80; Indels 61; Gaps 11;
 QY 18 YIGSQNASGRQRMRHNPVSGCGGCGCATCSQDN--GCLSKPRLFFALERIGMKQIGV 75
 DB 705 YFGDTAARRCR-----CHGCECTGCRSPQCLSCR-RGFI-----HHQSTNT 747

QY 76 CLSSPCSGYCTRYPDINKCTKADCDTCFNK-NFCTCKSGFYHLGKCLDNCPEGLE 134
 Db 748 CVTLCPCAGLYADESRL--CLRCHPSCKQKVCBGFSLARGSCIPDCPRTGY 805
 QY 135 ANNHTEMC-----VSIVHCEVS-----BNPSPCTYKKGKTCGFRKGTTET 175
 Db 806 FDESLIRGECCHTCTCVGSRBECIHCAKSPHQDKKCPACGE-----GF----- 853
 QY 176 VREIHPGSAKGNLCPPTNETRKTQVQRKCKQGER 211
 Db 854 -----YPEMPGL--PHKVRRCDCNCLSCBSSR 881

RESULT 10
 FSPO_XENLA STANDARD; PRT: 803 AA.
 AC P35447;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE F-spondin precursor.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=93376785; PubMed=8367492;
 RA Ruiz i Altaba A., Cox C., Jessell T.M., Klar A.;
 RT "Ectopic neural expression of a floor plate marker in frog embryos
 injected with the midline transcription factor Pintallavis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:8268-8272(1993).
 CC -1- FUNCTION: Promotes the attachment of spinal cord and sensory
 neuron cells and the outgrowth of neurites in vitro. May
 contribute to the growth and guidance of axons in both the spinal
 cord and the PNS.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: Expressed at high levels in the floor plate.
 CC -1- SIMILARITY: Contains 6 TSP type-1 domains.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

 CC EMBL; L09123; AAA19105.1; -;
 CC PIR; A47723; A47723.
 CC InterPro; IPR002861; Reeler.
 CC InterPro; IPR000884; TSP1.
 CC Pfam; PF02014; Reeler; 1.
 CC Pfam; PF00090; tsp1; 6.
 CC SMART; SM00209; TSP1; 6.
 CC PROSITE; PS00092; TSP1; 6.
 KW Glycoprotein; Signal; Repeat; Cell adhesion.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 803 P-SPONDIN.
 FT DOMAIN 436 489 TSP TYPE-1 1.
 FT DOMAIN 495 549 TSP TYPE-1 2.
 FT DOMAIN 552 605 TSP TYPE-1 3.
 FT DOMAIN 608 662 TSP TYPE-1 4.
 FT DOMAIN 664 717 TSP TYPE-1 5.
 FT DOMAIN 750 802 TSP TYPE-1 6.
 FT CARBOHYD 210 210 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 677 677 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 803 AA; 97070 MW; D3A54E329548AED9 CRC64;

Query Match 10.3%; Score 156; DB 1; Length 803;
 Best Local Similarity 24.6%; Pred. No. 0.00034;

Matches 58; Conservative 35; Mismatches 91; Indels 52; Gaps 12;
 QY 37 VSQCGGCGATCSDYNGCLSCFPLFPALERIGMKQIGVCLSSCPSS--GYGTRYPDINK 94
 Db 544 VNEBCEPSSCIVTWAEWBECS-----ATCRMGKGRHMIOTTPADGSMCKADATTEVEK 598
 QY 95 C--TKCK-----ADCDTCFNKPNCTCKSGFYHLGKCLDNCPEGLEANN-- 137
 Db 599 CMPECHTIPCVLSPWSEWSDCVTCGKTRTRQR---MLKSPSELGDCNBELELQVEK 655
 QY 138 -HTMECVSIVHCEVSEWNPSPCTKKGKTCGFRKTETVREIHPGSAKGNLCPPTNET 196
 Db 656 CMLPECP--ISCELTEWSYWSSEC---NKSOG--KGMIRTRMTWEPQFGGAVCPETVQR 708
 QY 197 RKTQVQRKCKQGERKGRKGRKPKNGESKEKRAIPDSKSLSSKEIPEORENKQ 252
 Db 709 KKCL--RKCK-----SSGNERRHLK-----DAREKRSSEKIKEDSDGQ 747

RESULT 11
 NTC2_MOUSE STANDARD; PRT: 2470 AA.
 ID NTC2_MOUSE
 AC O35516; Q06008; Q60941;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (MOTCH
 B).
 GN NOTCH2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Thymus;
 RA Hamada Y., Higuchi M., Tsujimoto Y.;
 RT "Complete amino acid sequence and multiform transcripts encoded by a
 single copy of mouse Notch2 gene.";
 RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 316-1518 FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
 RX MEDLINE=93178563; PubMed=8440332;
 RA Lardelli M., Lendahl U.;
 RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a
 wide variety of tissues.";
 RL Exp. Cell Res. 204:364-372(1993).
 RN [3]
 RP SEQUENCE OF 1765-2153 FROM N.A.
 RX MEDLINE=97075110; PubMed=8917536;
 RA Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,
 RA Martin D.I.;
 RT "Inhibition of granulocytic differentiation by mNotch1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).
 RN [4]
 RP FUNCTION.
 RX MEDLINE=99396706; PubMed=10393120;
 RA Hamada Y., Kadokawa Y., Okabe M., Ikawa M., Coleman J.R.,
 RA Tsujimoto Y.;
 RT "Mutation in ankyrin repeats of the mouse Notch2 gene induces early
 embryonic lethality.";
 RL Development 126:3415-3424(1999).
 RN [5]
 RP DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.
 RX MEDLINE=95333893; PubMed=7609614;
 RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;
 RT "Differential expression of Notch1 and Notch2 in developing and adult
 mouse brain.";
 RL Brain Res. Mol. Brain Res. 29:263-272(1995).
 RN [6]
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
 RX MEDLINE=21523956; PubMed=11518718;

RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.:
RT "Murine notch homologs (N1-4) undergo presenilin-dependent
RL proteolysis."; J. Biol. Chem. 276:40268-40273 (2001).
RN [7]
RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
RX MEDLINE=21374376; PubMed=11459941;
RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.:
RT "Conservation of the biochemical mechanisms of signal transduction
RL among mammalian Notch family members."; Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031 (2001).
CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (by similarity). May play an essential role in
CC postimplantation development, probably in some aspect of cell
CC specification and/or differentiation.
CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC terminal fragment N(EC) which are probably linked by disulfide
CC bonds.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=O35516-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O35516-2; Sequence=VSP_001405;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,
CC neuroepithelia, somites, optic vesicles and branchial arches, but
CC not heart.
CC -!- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,
CC the postnatal ependymal cells, and the choroid plexus throughout
CC embryonic and postnatal development.
CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane.
CC -!- PTM: Phosphorylated.
CC -!- SIMILARITY: Belongs to the NOTCH family.
CC -!- SIMILARITY: Contains 35 EGF-like domains.
CC -!- SIMILARITY: Contains 2 Lin/Notch repeats.
CC -!- SIMILARITY: Contains 6 ANK repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements/>
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL: D32210; BAA22094.1; -;
CC EMBL: X68279; CAA48340.1; -;
CC EMBL: U31881; AAC52924.1; -;
CC FTR: A49175; A49175.
CC HSSP: P16109; IFSB.
CC MGD: MGI:97364; Notch2.
CC GO: GO:0005887; C:integral to plasma membrane; IC.
CC GO: GO:0005515; P:protein binding; IPI.
CC GO: GO:0002011; P:morphogenesis of an epithelial sheet; IMP.
CC GO: GO:0007219; P:N signaling pathway; IC.

DR InterPro: IPR002110; ANK.
DR InterPro: IPR000152; Asx_hydroxyl_s.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001438; EGF_II.
DR InterPro: IPR006209; EGF_like.
DR InterPro: IPR002049; Laminin_EGF.
DR InterPro: IPR008297; Notch.
DR InterPro: IPR008000; Notch_dom.
DR Pfam: PF00023; ank; 6.
DR Pfam: PF00008; EGF; 34.
DR Pfam: PF00066; notch; 2.
DR PIRSF: PIRSF002279; Notch; 1.
DR PRINTS: PR00010; EGFBL00D.
DR PRINTS: PR00011; EGF_LAMININ.
DR PRINTS: PR01452; NOTCH.
DR SMART: SM00248; ANK; 6.
DR SMART: SM00179; EGF_CA; 23.
DR SMART: SM00004; NL; 3.
DR PROSITE: PS50297; ANK_REPEAT; 1.
DR PROSITE: PS50088; ANK_REPEAT; 4.
DR PROSITE: PS00010; ASX_HYDROXYL; 22.
DR PROSITE: PS00022; EGF_1; 33.
DR PROSITE: PS01186; EGF_2; 27.
DR PROSITE: PS00026; EGF_3; 35.
DR PROSITE: PS01187; EGF_CA; 22.
DR Receptor, Transcription regulation; Activator; Differentiation;
DR Developmental protein; Repeat; ANK repeat; EGF-like domain;
DR Transmembrane; Glycoprotein; Signal; Phosphorylation;
DR Alternative splicing.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 2470 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.
FT CHAIN 1665 2470 NOTCH EXTRACELLULAR TRUNCATION.
FT CHAIN 1697 2470 NOTCH INTRACELLULAR DOMAIN.
FT DOMAIN 26 1677 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1678 1698 POTENTIAL.
FT DOMAIN 1699 2470 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 26 63 EGF-LIKE 1.
FT DOMAIN 64 102 EGF-LIKE 2.
FT DOMAIN 105 143 EGF-LIKE 3.
FT DOMAIN 144 180 EGF-LIKE 4.
FT DOMAIN 182 219 EGF-LIKE 5.
FT DOMAIN 221 256 EGF-LIKE 6 (INCOMPLETE).
FT DOMAIN 258 294 EGF-LIKE 7.
FT DOMAIN 296 334 EGF-LIKE 8.
FT DOMAIN 336 372 EGF-LIKE 9.
FT DOMAIN 373 411 EGF-LIKE 10.
FT DOMAIN 413 452 EGF-LIKE 11.
FT DOMAIN 454 490 EGF-LIKE 12.
FT DOMAIN 492 528 EGF-LIKE 13.
FT DOMAIN 530 566 EGF-LIKE 14.
FT DOMAIN 568 603 EGF-LIKE 15.
FT DOMAIN 605 641 EGF-LIKE 16.
FT DOMAIN 643 678 EGF-LIKE 17.
FT DOMAIN 680 716 EGF-LIKE 18.
FT DOMAIN 718 753 EGF-LIKE 19.
FT DOMAIN 755 791 EGF-LIKE 20.
FT DOMAIN 793 829 EGF-LIKE 21.
FT DOMAIN 831 869 EGF-LIKE 22.
FT DOMAIN 871 907 EGF-LIKE 23.
FT DOMAIN 909 945 EGF-LIKE 24.
FT DOMAIN 947 983 EGF-LIKE 25.
FT DOMAIN 985 1021 EGF-LIKE 26.
FT DOMAIN 1023 1059 EGF-LIKE 27.
FT DOMAIN 1061 1097 EGF-LIKE 28.
FT DOMAIN 1099 1145 EGF-LIKE 29.
FT DOMAIN 1147 1183 EGF-LIKE 30.
FT DOMAIN 1185 1221 EGF-LIKE 31.
FT DOMAIN 1223 1260 EGF-LIKE 32.
FT DOMAIN 1262 1300 EGF-LIKE 33.
FT DOMAIN 1302 1345 EGF-LIKE 34.
FT DOMAIN 1372 1410 EGF-LIKE 35.
FT REPEAT 1418 1454 LIN/NOTCH 1.

FT DISULFID 47 BY SIMILARITY.
FT DISULFID 62 BY SIMILARITY.
FT DISULFID 68 BY SIMILARITY.
FT DISULFID 89 BY SIMILARITY.
FT DISULFID 106 BY SIMILARITY.
FT DISULFID 111 BY SIMILARITY.
FT DISULFID 128 BY SIMILARITY.
FT DISULFID 130 BY SIMILARITY.
FT DISULFID 145 BY SIMILARITY.
FT DISULFID 150 BY SIMILARITY.
FT DISULFID 167 BY SIMILARITY.
FT DISULFID 176 BY SIMILARITY.
FT DISULFID 183 BY SIMILARITY.
FT DISULFID 188 BY SIMILARITY.
FT DISULFID 205 BY SIMILARITY.
FT DISULFID 221 BY SIMILARITY.
FT DISULFID 226 BY SIMILARITY.
FT DISULFID 244 BY SIMILARITY.
FT DISULFID 260 BY SIMILARITY.
FT DISULFID 265 BY SIMILARITY.
FT DISULFID 282 BY SIMILARITY.
FT DISULFID 298 BY SIMILARITY.
FT DISULFID 305 BY SIMILARITY.
FT DISULFID 322 BY SIMILARITY.
FT DISULFID 338 BY SIMILARITY.
FT DISULFID 343 BY SIMILARITY.
FT DISULFID 360 BY SIMILARITY.
FT DISULFID 375 BY SIMILARITY.
FT DISULFID 380 BY SIMILARITY.
FT DISULFID 399 BY SIMILARITY.
FT DISULFID 415 BY SIMILARITY.
FT DISULFID 422 BY SIMILARITY.
FT DISULFID 439 BY SIMILARITY.
FT DISULFID 455 BY SIMILARITY.
FT DISULFID 460 BY SIMILARITY.
FT DISULFID 477 BY SIMILARITY.
FT DISULFID 493 BY SIMILARITY.
FT DISULFID 498 BY SIMILARITY.
FT DISULFID 513 BY SIMILARITY.
FT DISULFID 515 BY SIMILARITY.
FT DISULFID 531 BY SIMILARITY.
FT DISULFID 536 BY SIMILARITY.
FT DISULFID 553 BY SIMILARITY.
FT DISULFID 569 BY SIMILARITY.
FT DISULFID 574 BY SIMILARITY.
FT DISULFID 590 BY SIMILARITY.
FT DISULFID 606 BY SIMILARITY.
FT DISULFID 611 BY SIMILARITY.
FT DISULFID 628 BY SIMILARITY.
FT DISULFID 644 BY SIMILARITY.
FT DISULFID 649 BY SIMILARITY.
FT DISULFID 665 BY SIMILARITY.
FT DISULFID 681 BY SIMILARITY.
FT DISULFID 686 BY SIMILARITY.
FT DISULFID 703 BY SIMILARITY.
FT DISULFID 719 BY SIMILARITY.
FT DISULFID 724 BY SIMILARITY.
FT DISULFID 740 BY SIMILARITY.
FT DISULFID 756 BY SIMILARITY.
FT DISULFID 761 BY SIMILARITY.
FT DISULFID 778 BY SIMILARITY.
FT DISULFID 794 BY SIMILARITY.
FT DISULFID 799 BY SIMILARITY.
FT DISULFID 816 BY SIMILARITY.
FT DISULFID 832 BY SIMILARITY.
FT DISULFID 837 BY SIMILARITY.
FT DISULFID 854 BY SIMILARITY.
FT DISULFID 865 BY SIMILARITY.
FT DISULFID 872 BY SIMILARITY.
FT DISULFID 877 BY SIMILARITY.
FT DISULFID 894 BY SIMILARITY.
FT DISULFID 910 BY SIMILARITY.
FT DISULFID 915 BY SIMILARITY.
FT DISULFID 932 BY SIMILARITY.
FT DISULFID 986 BY SIMILARITY.
FT DISULFID 991 BY SIMILARITY.
FT DISULFID 1008 BY SIMILARITY.

FT DISULFID 1024 BY SIMILARITY.
FT DISULFID 1029 BY SIMILARITY.
FT DISULFID 1046 BY SIMILARITY.
FT DISULFID 1062 BY SIMILARITY.
FT DISULFID 1067 BY SIMILARITY.
FT DISULFID 1084 BY SIMILARITY.
FT DISULFID 1100 BY SIMILARITY.
FT DISULFID 1115 BY SIMILARITY.

Query Match 9.8%; Score 148.5; DB 1; Length 2524;
Best Local Similarity 22.5%; Pred. No. 0.0035;
Matches 73; Conservative 32; Mismatches 102; Indels 117; Gaps 19;

Qy 33 MHPVSOQGGCATCSYNGCLSKPLPALERIGH-----KQIGV 75
Db 869 MNECVNRPCKNG-ATQONTNGSYKNCCKPGYGRNCEIDDDCQNPCHNGSCSDGINM 927
Qy 76 CLSSCPGYGTR-YPDINKCTK--CK--ABCDTCFNKNFTCKSGFY-LHL-----122
Db 928 PFCNCPAGFRGPKCEEDINECASNPCKNGANGCTCVNSVTCT-CQPGFSGIHCSNTPDC 986
Qy 123 -----GKCLD-----NCPBGLAN--NHTM-BCVSI-----145
Db 987 TESSCFNGGTCIDGINTPTCQCPGFTGSCYCHDINECDKPLANGTQDSYGYTKCTC 1046
Qy 146 -----VHCE-VSEWNPMSPTCKGK-----TCGFKRG-----TETRVRE 178
Db 1047 PQGTGLNQNVLVRWCDSPPCKNGKCKWQTNFRCCKSGMTGYCDVPSVCEVAAKQ 1106
Qy 179 ---IIQHPSAKGNLCPTNETRKTCTVQRKCKQKGRGKGRK-RPNK-----225
Db 1107 QGVDIHLCRNSGMCVDTGNTHFC-----RCQAGYTGSCYCBQVDECSNPQNGATCTD 1161

Qy 226 ---GESKEALPDSKSLSSKEIPE 246
Db 1162 YLGYSCECVAGYHGVNCSBINE 1185

RESULT 14
NTC2 RAT
ID NTC2 RAT STANDARD; PRT; 2471 AA.
AC Q9QW30;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DB Neurogenic locus notch homolog protein 2 precursor (Notch 2).
GN NOTCH2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RS TISSUE=Brain;
RX MEDLINE=93202015; PubMed=1295745;
RA Weinmaster G., Roberts V.J., Lemke G.;
RT "Notch2: a second mammalian Notch gene.";
RL Development 116:931-941(1992).
RN [2]
RS TISSUE SPECIFICITY.
RX MEDLINE=2131789; PubMed=11438922;
RA Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
functional roles for the Notch-DSL signaling system during brain
development.";
RL J. Comp. Neurol. 436:167-181(2001).
CC -I- FUNCTION: Functions as a receptor for membrane-bound ligands
Jagged1, Jagged2 and Deltal to regulate cell-fate determination.
Upon ligand activation through the released notch intracellular
domain (NICD) it forms a transcriptional activator complex with
RBP-J kappa and activates genes of the enhancer of split locus.
Affects the implementation of differentiation, proliferation and
apoptotic programs. May play an essential role in postimplantation

development, probably in some aspect of cell specification and/or differentiation (By similarity).

-!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-terminal fragment N(EC) which are probably linked by disulfide bonds (By similarity).

-!- SUBCELLULAR LOCATION: Type I membrane protein. Following proteolytical processing NiCD is translocated to the nucleus.

-!- TISSUE SPECIFICITY: Highly expressed in the spleen and choroid plexus in the brain. Expressed in postnatal central nervous system (CNS) germinal zones and, in early postnatal life, within numerous cells throughout the CNS. It is more highly localized to ventricular germinal zones. Also found in the heart, liver and kidney.

-!- DEVELOPMENTAL STAGE: Expressed in the brain during E14 and E17.

-!- PTM: Synthesized in the endoplasmic reticulum as an inactive form which is proteolytically cleaved by a furin-like convertase in the trans-Golgi network before it reaches the plasma membrane to yield an active, ligand-accessible form. Cleavage results in a C-terminal fragment N(TM) and a N-terminal fragment N(EC). Following ligand binding, it is cleaved by TNF-alpha converting enzyme (TACE) to yield a membrane-associated intermediate fragment called notch extracellular truncation (NEXT). This fragment is then cleaved by presenilin dependent gamma-secretase to release a notch-derived peptide containing the intracellular domain (NiCD) from the membrane (By similarity).

-!- PTM: Phosphorylated (By similarity).

-!- SIMILARITY: Belongs to the NOTCH family.

-!- SIMILARITY: Contains 35 EGF-like domains.

-!- SIMILARITY: Contains 2 lin/Notch repeats.

-!- SIMILARITY: Contains 6 ANK repeats.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/> or send an email to license@sib-sib.ch).

EMBL: M93661; AAK13558.1; --
 PIR: A49128; A49128.
 HSP: P00743; ICCP.
 InterPro: IPR002110; ANK.
 InterPro: IPR000152; Asx hydroxyl S.
 InterPro: IPR000742; EGF 2.
 InterPro: IPR001881; EGF Ca.
 InterPro: IPR001436; EGF II.
 InterPro: IPR006209; EGF like.
 InterPro: IPR002049; Laminin_EGF.
 InterPro: IPR008297; Notch.
 InterPro: IPR008000; Notch_dom.
 Pfam: PF00023; ank; 6.
 Pfam: PF00008; EGF; 35.
 Pfam: PF00066; notch; 2.
 PRINTS: PIRSF002279; Notch; 1.
 PRINTS: PR00010; EGFBL00D.
 PRINTS: PR00011; EGFLAMININ.
 PRINTS: PR01452; NOTCH.
 SMART: SM00248; ANK; 6.
 SMART: SM00179; EGF CA; 24.
 SMART: SM00004; NL; 2.
 PROSITE: PS50297; ANK REP REGION; 1.
 PROSITE: PS50088; ANK_REPEAT; 4.
 PROSITE: PS00010; ASX_HYDROXYL; 22.
 PROSITE: PS00022; EGF_1; 34.
 PROSITE: PS01186; EGF_2; 26.
 PROSITE: PS50026; EGF_3; 35.
 PROSITE: PS01187; EGF CA; 22.
 Receptor; Transcription regulation; Activator; Differentiation;
 Developmental protein; Repeat; ANK repeat; EGF-like domain;
 Transmembrane; Glycoprotein; Signal; Phosphorylation.
 SIGNAL 1 25
 CHAIN 26 2471 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.

CHAIN	1666	2471	NOTCH EXTRACELLULAR TRUNCATION (BY SIMILARITY).
FT	1666	2471	NOTCH EXTRACELLULAR TRUNCATION (BY SIMILARITY).
FT	1697	2471	NOTCH INTRACELLULAR DOMAIN (BY SIMILARITY).
FT	26	1677	EXTRACELLULAR (POTENTIAL).
FT	1678	1698	CYTOPLASMIC (POTENTIAL).
FT	1699	2471	EGF-LIKE 1.
FT	26	63	EGF-LIKE 2.
FT	105	143	EGF-LIKE 3.
FT	144	180	EGF-LIKE 4.
FT	182	219	EGF-LIKE 5.
FT	221	258	EGF-LIKE 6.
FT	260	296	EGF-LIKE 7.
FT	298	336	EGF-LIKE 8.
FT	338	374	EGF-LIKE 9.
FT	375	413	EGF-LIKE 10.
FT	415	454	EGF-LIKE 11.
FT	456	492	EGF-LIKE 12.
FT	494	530	EGF-LIKE 13.
FT	532	568	EGF-LIKE 14.
FT	570	605	EGF-LIKE 15.
FT	607	643	EGF-LIKE 16.
FT	645	680	EGF-LIKE 17.
FT	682	718	EGF-LIKE 18.
FT	720	755	EGF-LIKE 19.
FT	757	793	EGF-LIKE 20.
FT	795	831	EGF-LIKE 21.
FT	833	871	EGF-LIKE 22.
FT	873	909	EGF-LIKE 23.
FT	911	947	EGF-LIKE 24.
FT	949	985	EGF-LIKE 25.
FT	987	1023	EGF-LIKE 26.
FT	1025	1061	EGF-LIKE 27.
FT	1063	1099	EGF-LIKE 28.
FT	1101	1147	EGF-LIKE 29.
FT	1149	1185	EGF-LIKE 30.
FT	1187	1223	EGF-LIKE 31.
FT	1225	1262	EGF-LIKE 32.
FT	1264	1302	EGF-LIKE 33.
FT	1304	1343	EGF-LIKE 34.
FT	1374	1412	EGF-LIKE 35.
FT	1645	1682	POLY-ALA.
FT	1994	1997	POLY-SER.
FT	2426	2429	POLY-GLY.
FT	2446	2451	LIN/NOTCH 1.
FT	1420	1456	LIN/NOTCH 2.
FT	1503	1535	ANK 1.
FT	1827	1871	ANK 2.
FT	1876	1905	ANK 3.
FT	1909	1939	ANK 4.
FT	1943	1972	ANK 5.
FT	1976	2005	ANK 6.
FT	2009	2038	BY SIMILARITY.
FT	28	41	BY SIMILARITY.
FT	35	51	BY SIMILARITY.
FT	53	62	BY SIMILARITY.
FT	68	79	BY SIMILARITY.
FT	73	90	BY SIMILARITY.
FT	92	101	BY SIMILARITY.
FT	109	121	BY SIMILARITY.
FT	115	131	BY SIMILARITY.
FT	133	142	BY SIMILARITY.
FT	148	159	BY SIMILARITY.
FT	153	168	BY SIMILARITY.
FT	170	179	BY SIMILARITY.
FT	186	198	BY SIMILARITY.
FT	192	207	BY SIMILARITY.
FT	209	218	BY SIMILARITY.
FT	225	236	BY SIMILARITY.
FT	230	245	BY SIMILARITY.
FT	238	257	BY SIMILARITY.
FT	264	275	BY SIMILARITY.

```

FT DISULFID 269 284 BY SIMILARITY.
FT DISULFID 286 295 BY SIMILARITY.
FT DISULFID 302 315 BY SIMILARITY.
FT DISULFID 309 324 BY SIMILARITY.
FT DISULFID 326 335 BY SIMILARITY.
FT DISULFID 342 353 BY SIMILARITY.
FT DISULFID 347 362 BY SIMILARITY.
FT DISULFID 364 373 BY SIMILARITY.
FT DISULFID 379 390 BY SIMILARITY.
FT DISULFID 384 401 BY SIMILARITY.
FT DISULFID 403 412 BY SIMILARITY.
FT DISULFID 419 433 BY SIMILARITY.
FT DISULFID 427 442 BY SIMILARITY.
FT DISULFID 444 453 BY SIMILARITY.
FT DISULFID 460 471 BY SIMILARITY.
FT DISULFID 465 480 BY SIMILARITY.
FT DISULFID 482 491 BY SIMILARITY.
FT DISULFID 498 509 BY SIMILARITY.
FT DISULFID 503 518 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 536 547 BY SIMILARITY.

Query Match 9.8%; Score 147.5; DB 1; Length 2471;
Best Local Similarity 24.8%; Pred. NO. 0.004;
Matches 54; Conservative 23; Mismatches 52; Indels 89; Gaps 13;

Qy 30 QRMRHPNVSCQCGCATCSYNGCLSCPKPLPALERIGMKQIGVCLSSCPSCGYGTRY 89
Db 947 QTDVNECLSEPCXNG-GTCDDVNSYTC-----TCPAGFHGVC 984
Qy 90 P-DINKCTCKKADCTCFNKFC-----TKCKGFFVLHGK 124
Db 985 ENNIDECTE-----SSFCNGTCVDGINSFSCLPVGTGPFCLHDINECSSNCLNGST 1039
Qy 125 CLD-----NCPGLRANNHTECVSIHV-CRVSEMPNPSCTKKGKTCGFKRGTRV 176
Db 1040 CVDGLGTVRCTPLGYTGKN---CQTLVNLG-----SP-SPCKNGK-TCA----- 1079
Qy 177 REIQHPSAKNGLCPPTNETKCTVQKCK-----QKG 209
Db 1080 ----QEKARPRCLCPGCGMGAYCDLVNYSCKAALQKG 1113

RESULT 15
BL14_CABEL
ID BL14_CABEL STANDARD; PRT; 943 AA.
AC P51559; O44762; O44763; O44764; O44765; O44766;
DT 01-OCT-1996 (Rel. 34, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endoprotease bli-4 precursor (EC 3.4.21.-) (Blisterase) (Blistered
cuticle protein 4).
GN BLI-4 OR KPC-4 OR K04F10.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RN SEQUENCE FROM N.A. (ISOFORMS A; B; C AND D), AND FUNCTION.
RP STRAIN=Bristol N2;
RC MEDLINE=95293228; PubMed=7774813;
RX Thacker C., Peters K.W., Grayko M., Rose A.M.;
RT "The bli-4 locus of Caenorhabditis elegans encodes structurally
distinct kex2/subtilisin-like endoproteases essential for early
development and adult morphology.";
RL Genes Dev. 9:956-971(1995).
RN [2]
RN SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RP STRAIN=Bristol N2;
RC Latreille P., Wamsley P.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP MUTAGENESIS.

```

```

RX MEDLINE=20363723; PubMed=10903434;
RA Thacker C., Grayko M., Rose A.M.;
RT "Mutational analysis of bli-4/kpc-4 reveals critical residues required
for proprotein convertase function in C. elegans.";
RL Gene 252:15-25(2000).
CC -!- FUNCTION: The kex2/subtilisin-like proteinase activity of this
enzyme is required for the normal production of adult cuticle.
There is functional redundancy between the isoforms. Vital for
embryonic and larval development.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=5;
CC Name=D; Synonyms=d;
CC IsoId=P51559-1; Sequence=Displayed;
CC Name=A; Synonyms=a;
CC IsoId=P51559-2; Sequence=VSP_005416, VSP_005419;
CC Name=B; Synonyms=b;
CC IsoId=P51559-3; Sequence=VSP_005418, VSP_005421;
CC Name=C; Synonyms=c;
CC IsoId=P51559-4; Sequence=VSP_005422, VSP_005423;
CC Name=E;
CC IsoId=P51559-5; Sequence=VSP_005417, VSP_005420;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: In larvae and adults, expressed in all
hypodermal cells, vulva and ventral nerve cords.
CC -!- DEVELOPMENTAL STAGE: Expression starts at embryo two-fold stage
through to adults.
CC -!- SIMILARITY: Belongs to peptidase family S8. Furin subfamily.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
frameshift in position 651.
CC -!- CAUTION: Ref.2 (AAB96754 and AAB96757) sequence differs from that
shown due to erroneous gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
EMBL; L29438; AAA98750.1; ALT_FRAME.
EMBL; L29439; AAA98751.1; ALT_FRAME.
EMBL; L29440; AAA98752.1; ALT_FRAME.
EMBL; AF039719; AAB96753.1; -; -;
EMBL; AF039719; AAB96754.1; ALT_SEQ.
EMBL; AF039719; AAB96755.1; -; -;
EMBL; AF039719; AAB96756.1; -; -;
EMBL; AF039719; AAB96757.1; ALT_SEQ.
MEROPS; S08.031; -; -;
WormPep; K04F10.4a; CE11728.
WormPep; K04F10.4b; CE11730.
WormPep; K04F10.4c; CE11732.
WormPep; K04F10.4d; CE11734.
WormPep; K04F10.4e; CE11736.
GO; GO:0016021; C: integral to membrane; NAS.
GO; GO:0004252; P: serine-type endopeptidase activity; NAS.
GO; GO:0007552; P: cuticle biosynthesis (sensu invertebrata); IMP.
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow fac recep.
DR InterPro; IPR002029; Peptidase S8.
DR InterPro; IPR002884; Peptidase S8B.
DR InterPro; IPR009020; Protease Inhib.
DR Pfam; PF01483; P_protein; 1.
DR Pfam; PF00082; Peptidase S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR ProDom; PD000717; P domain; 1.
DR SMART; SM00261; FU_3.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Calcium-binding; Zymogen;
KW Transmembrane; Signal; Alternative splicing.
FT SIGNAL 1 20

```



```
FT PROPEP      21      116      BY SIMILARITY.
FT CHAIN       117      943      ENDOPEPTASE BLI-4.
FT ACT_SITE    202      202      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE    241      241      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE    415      415      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT TRANSMEM    417      437      POTENTIAL.
FT DOMAIN      438      684      CYTOPLASMIC (POTENTIAL).
FT DOMAIN      713      857      CYS-RICH.
FT CARBOHYD    195      195      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC    658      670      VERSARSPFDLT -> ILITIALHLVNA (in
FT                                     isoform A).
FT                                     /FTid=VSP_005416.
FT VARSPLIC    658      699      VERSARSPFDLTSGWKLSCDECNCGCTESSATSCPAYKH
FT                                     L -> LCFNENSOPTCEVCLVCEPGVIINFQIEKLKST.
FT                                     KCLIFS (in isoform E).
FT                                     /FTid=VSP_005417.
FT VARSPLIC    658      730      VERSARSPFDLTSGWKLSCDECNCGCTESSATSCPAYKH
FT                                     LITQLRNKGSGFKVQKCDYTYLDGDKCM -> GDEVV
FT                                     ERIKNHWEVTLEESSHNWEHREHKSLOBLNLSRTHSPL
FT                                     YSTKFKQPIFLIILVCIFDAIHRQFAV (in isoform
FT                                     B).
FT                                     /FTid=VSP_005418.
FT                                     Missing (in isoform A).
FT VARSPLIC    671      943      /FTid=VSP_005419.
FT VARSPLIC    700      943      Missing (in isoform E).
FT VARSPLIC    731      943      /FTid=VSP_005420.
FT VARSPLIC    780      828      Missing (in isoform B).
FT                                     /FTid=VSP_005421.
FT VARSPLIC    829      943      ESNLVQAKCIWRKDLCDGYYINAVGKCDLDCSCTCTAP
FT                                     GPMSCERC -> AENFDPCAKNNEGRDITVFKFKPSAP
FT                                     KDVCNLKLVLDPPFISLLP (in isoform C).
FT                                     /FTid=VSP_005422.
FT VARSPLIC    829      943      Missing (in isoform C).
FT                                     /FTid=VSP_005423.
FT CONFLICT    153      153      A -> R (IN REF. 2).
FT CONFLICT    346      346      S -> R (IN REF. 2).
FT SEQUENCE    943 AA; 103146 MW; 401E09B6C46AD7E CRC64;

Query Match      9.6%; Score 145; DB 1; Length 943;
Best Local Similarity 25.6%; Pred. No. 0.0024;
Matches 51; Conservative 21; Mismatches 77; Indels 50; Gaps 11;

QY 31 RRMHPNVSQ-----COGSCATCSDYNGCLSCPKRLFFALERIGMKQIGVCLSSCPFG 83
Db 663 RSSFPDLTSGWKLSCDECNCGCTESSATSCPAYK-HLTQTLRNKGGSGP-KCVQKDDT 720
QY 84 YGTRYPDINKTKCKADCTCPNKNFCTKCKSGFYL-----HL--GKCLDNCPEGLE 134
Db 721 Y-----YLDGDKCKMCSHCHCTCTKAEVCETCPGSLILLIDVDMPHYDHGKCVESCPGLV 776
QY 135 ANNHTECVSIHCEVSEWNPSPCTKKGKTCGPKGTETVRELIQHPSAKGNLCPPTN 194
Db 777 ADYES-----NLVQAK-----CIWRKDLCG-----DGYINAVGKCDLDCSSC 814
QY 195 ETRKCT----VORKKCKOG 209
Db 815 ET--CTAPGPMSECKSKG 831
```

Search completed: June 29, 2004, 17:00:18
Job time : 9.25243 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 16:55:24 ; Search time 34.6602 Seconds

(without alignments)
2476.067 Million cell updates/sec

Title: US-09-894-912A-34

Perfect score: 1510

Sequence: 1 MHLRLISWLFILNFMFYIG.....QQKKRKVDKQKSVSVSTVH 272

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:**

1: sp_archaea:**

2: sp_bacteria:**

3: sp_fungi:**

4: sp_human:**

5: sp_invertebrate:**

6: sp_mammal:**

7: sp_mhc:**

8: sp_organelle:**

9: sp_phage:**

10: sp_plant:**

11: sp_rodent:**

12: sp_virus:**

13: sp Vertebrate:**

14: sp Unclassified:**

15: sp_rvirus:**

16: sp_bacteriap:**

17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1510	100.0	272	4 Q9BX4	Q9bx4 homo sapien
2	1457	96.5	292	4 Q96K87	Q96k87 homo sapien
3	1158	76.7	217	11 Q9C3B2	Q9c3b2 mus musculus
4	1069	70.8	224	11 Q8BVW2	Q8bv2 mus musculus
5	644	42.6	265	11 Q92132	Q92132 mus musculus
6	626.5	41.5	236	4 Q8N7L5	Q8n7l5 homo sapien
7	578.5	38.3	243	11 Q8BFU0	Q8bfu0 mus musculus
8	472	31.3	224	4 Q9UGB2	Q9ugb2 homo sapien
9	459.5	30.4	176	4 Q8N6X6	Q8n6x6 homo sapien
10	419.5	27.8	152	11 Q7TPX3	Q7tpx3 mus musculus
11	263.5	18.8	138	11 Q8BJ73	Q8bj73 mus musculus
12	192	12.7	1299	5 Q26489	Q26489 spodoptera
13	179.5	11.9	913	13 Q8A1I8	Q8ay18 rana escul
14	177	11.7	1101	5 Q964D2	Q964d2 entamoeba h
15	176.5	11.7	1376	5 Q8S2S2	Q8szs2 drosophila
16	176	11.7	1074	5 Q964D1	Q964d1 entamoeba h

17	172	11.4	915	11	Q91VK0	Q91vk0 mus musculu
18	167.5	11.1	296	11	Q35171	Q35171 mus musculu
19	167.5	11.1	826	11	Q8CF22	Q8cf22 mus musculu
20	167.5	11.1	932	11	Q62030	Q62030 mus musculu
21	163	10.8	3869	5	Q86PQ3	Q86pq3 cryptospori
22	162	10.7	913	4	Q96BP4	Q96bp4 homo sapien
23	161	10.7	440	5	O18003	O18003 caenorhabdi
24	159	10.5	503	5	Q9U018	Q9u018 giardia lam
25	157.5	10.4	803	13	Q42114	Q42114 brachydanio
26	156	10.3	898	5	O76822	O76822 brachioosto
27	155	10.3	808	13	Q42113	Q42113 brachydanio
28	153.5	10.2	1362	13	Q9PV24	Q9pv24 xenopus lae
29	151	10.0	238	5	O76510	O76510 cryptospori
30	149.5	9.9	4007	4	Q86XX4	Q86xx4 homo sapien
31	149	9.9	1371	11	Q9QVW4	Q9qv4 rattus sp.
32	148	9.8	548	5	Q9GQ45	Q9gq45 giardia lam
33	148	9.8	4010	11	Q80T14	Q80t14 mus musculu
34	147.5	9.8	802	13	Q9W770	Q9w770 gallus gall
35	146.5	9.7	365	11	Q924Y6	Q924y6 rattus norv
36	146.5	9.7	807	4	Q9HC86	Q9hcb6 homo sapien
37	145	9.6	724	4	O94862	O94862 homo sapien
38	145	9.6	807	4	Q8NCD7	Q8ncd7 homo sapien
39	144.5	9.6	660	5	Q23832	Q23832 cryptospori
40	144.5	9.6	2189	5	Q9BI05	Q9bi05 eimeria ten
41	144	9.5	807	6	Q9GLX9	Q9glx9 bos taurus
42	143.5	9.5	220	11	Q99KR2	Q99kr2 mus musculu
43	143.5	9.5	402	11	Q8K2Q8	Q8k2q8 mus musculu
44	143.5	9.5	807	11	Q8VCC9	Q8vcc9 mus musculu
45	143	9.5	435	5	Q9GQ41	Q9gq41 giardia lam

ALIGNMENTS

RESULT 1

Q9BX4

ID Q9BX4 PRELIMINARY; PRT; 272 AA.

AC Q9BX4;

DT 01-JUN-2001 (TRENBLrel. 17, Created)

DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Thrombospondin.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Mao Y., Xie Y., Zhou Z., Zhao W., Zhao S., Wang W., Huang Y., Wang S.,

RA Tang R., Chen X., Wu C.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Placenta;

RA Strausberg R.;

RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF251057; AAK34947.1; --

DR EMBL; BC022367; AAH22367.1; --

DR Genew; HGNC:20866; THSD2.

DR InterPro; IPR006212; Furin repeat.

DR InterPro; IPR009030; Grow_fac_recep.

DR InterPro; IPR000884; TSP1.

DR Pfam; PF00090; tsp_1; 1.

DR SMART; SM00261; FU; 2.

DR SMART; SM00209; TSP1; 1.

DR PROSITE; PS50092; TSP1; 1.

SQ SEQUENCE 272 AA; 3028 MW; CACAC6B78189 CRC64;

Query Match 100.0%; Score 1510; DB 4; Length 272;

Best Local Similarity 100.0%; Pred. No. 3.5e-130; Indels 0; Gaps 0;

Matches 272; Conservative 0; Mismatches 0;

QY 1 MHLRLISWLFILNFMFYIGSQNASRGRQRMRHFNVSQGGCGCATCSYNGCLSKR 60

Db 1 MHRLISWLFILINFMFYEYIGSNASRRGRQRMRHNPVSGCGGCATCSYNGCLSCKPR 60
 QY 61 LFPALERIGMKQIGVCLSSCPGSGYGYTRYDPDINKTKCKADCDTCFNKNFCTCKSGFFYL 120
 Db 61 LFPALERIGMKQIGVCLSSCPGSGYGYTRYDPDINKTKCKADCDTCFNKNFCTCKSGFFYL 120
 QY 121 HLGKCLDNCPEGLANNHTMECVSIHVCEVSEWNPSPCTCKGKTCGFKRGTTETRVREII 180
 Db 121 HLGKCLDNCPEGLANNHTMECVSIHVCEVSEWNPSPCTCKGKTCGFKRGTTETRVREII 180
 QY 181 QHPSAKGNLCPPPTNETRKTCTVQRKKQKGERGKGRERKRPKNKGSKEALPDSKSLES 240
 Db 181 QHPSAKGNLCPPPTNETRKTCTVQRKKQKGERGKGRERKRPKNKGSKEALPDSKSLES 240
 QY 241 SKEIPQRENKQKKRKVKQDKQKSVSVSTVH 272
 Db 241 SKEIPQRENKQKKRKVKQDKQKSVSVSTVH 272

RESULT 2

Q96K87 PRELIMINARY; PRT; 292 AA.
 AC Q96K87;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein FLJ14440.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
 RA Arica M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
 RA Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y., Oshima A.;
 RT "NEO human cDNA sequencing project."
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK027346; BAB55051.1; --
 DR InterPro: IPR006212; Furin repeat.
 DR InterPro: IPR009030; Grow_fac_recep.
 DR InterPro: IPR000884; TSP1.
 DR Pfam: PF00090; tsp.1; 1.
 DR SMART; SM00261; FU; 2.
 DR SMART; SM00209; TSP1; 1.
 DR PROSITE; PS50092; TSP1; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 292 AA; 33243 MW; 01E2774AC3D4A6F8 CRC64;

Query Match 96.5%; Score 1457; DB 4; Length 292;
 Best Local Similarity 99.2%; Pred. No. 2.7e-125;
 Matches 263; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MHRLISWLFILINFMFYEYIGSNASRRGRQRMRHNPVSGCGGCATCSYNGCLSCKPR 60
 Db 1 MHRLISWLFILINFMFYEYIGSNASRRGRQRMRHNPVSGCGGCATCSYNGCLSCKPR 60
 QY 61 LFPALERIGMKQIGVCLSSCPGSGYGYTRYDPDINKTKCKADCDTCFNKNFCTCKSGFFYL 120
 Db 61 LFPALERIGMKQIGVCLSSCPGSGYGYTRYDPDINKTKCKADCDTCFNKNFCTCKSGFFYL 120
 QY 121 HLGKCLDNCPEGLANNHTMECVSIHVCEVSEWNPSPCTCKGKTCGFKRGTTETRVREII 180
 Db 121 HLGKCLDNCPEGLANNHTMECVSIHVCEVSEWNPSPCTCKGKTCGFKRGTTETRVREII 180
 QY 181 QHPSAKGNLCPPPTNETRKTCTVQRKKQKGERGKGRERKRPKNKGSKEALPDSKSLES 240
 Db 181 QHPSAKGNLCPPPTNETRKTCTVQRKKQKGERGKGRERKRPKNKGSKEALPDSKSLES 240

QY 241 SKEIPQRENKQKKRKVKQDKQKSVSVSTVH 265
 Db 241 SKEIPQRENKQKKRKVKQDKQKSVSVSTVH 265
 RESULT 3
 Q9CSB2 PRELIMINARY; PRT; 217 AA.
 ID Q9CSB2;
 AC Q9CSB2;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE THSD2 OR 2810459H04RIK.
 GN Mus musculus (Mouse).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Embryo;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690 (2001).
 DR EMBL; AK013366; BAB28811.1; --
 DR MGI; MGI:1920030; Thsd2.
 DR InterPro: IPR006212; Furin repeat.
 DR InterPro: IPR009030; Grow_fac_recep.
 DR InterPro: IPR000884; TSP1.
 DR Pfam: PF00090; tsp.1; 1.
 DR SMART; SM00261; FU; 2.
 DR SMART; SM00209; TSP1; 1.
 DR PROSITE; PS50092; TSP1; 1.
 FT NON_TER 217 217
 SQ SEQUENCE 217 AA; 24304 MW; 0DCE938E9FB3FBF7 CRC64;

Query Match 76.7%; Score 1158; DB 11; Length 217;
 Best Local Similarity 93.1%; Pred. No. 4.5e-98;
 Matches 202; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 1 MHRLISWLFILINFMFYEYIGSNASRRGRQRMRHNPVSGCGGCATCSYNGCLSCKPR 60
 Db 1 MHRLISWLFILINFMFYEYIGSNASRRGRQRMRHNPVSGCGGCATCSYNGCLSCKPR 60
 QY 61 LFPALERIGMKQIGVCLSSCPGSGYGYTRYDPDINKTKCKADCDTCFNKNFCTCKSGFFYL 120
 Db 61 LFPALERIGMKQIGVCLSSCPGSGYGYTRYDPDINKTKCKADCDTCFNKNFCTCKSGFFYL 120
 QY 121 HLGKCLDNCPEGLANNHTMECVSIHVCEVSEWNPSPCTCKGKTCGFKRGTTETRVREII 180
 Db 121 HLGKCLDNCPEGLANNHTMECVSIHVCEVSEWNPSPCTCKGKTCGFKRGTTETRVREII 180
 QY 181 QHPSAKGNLCPPPTNETRKTCTVQRKKQKGERGKGRERKRPKNKGSKEALPDSKSLES 240
 Db 181 QHPSAKGNLCPPPTNETRKTCTVQRKKQKGERGKGRERKRPKNKGSKEALPDSKSLES 240

```

RESULT 4
ID Q8BVW2 PRELIMINARY; PRT; 224 AA.
AC Q8BVW2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Thrombospondin homolog.
GN THSD2 OR 2810459H04RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=CS7BL/6J; TISSUE=SKIN;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK076308; BAC36296.1; --
DR MGD; MGI:1920030; Thsd2.
DR InterPro; IPR008212; Furin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR SMART; SM00261; FU; 2.
SQ SEQUENCE 224 AA; 25398 MW; BC13B083497CFEE3 CRC64;

Query Match 70.8%; Score 1069; DB 11; Length 224;
Best Local Similarity 84.7%; Pred. No. 6.5e-90;
Matches 188; Conservative 9; Mismatches 25; Indels 0; Gaps 0;

Qy 1 MHLRLISWLFILNFMFYIGSNASRGRRQRRMHPNVSQCGGCATCSDYNGCLSKPR 60
Db 1 MHLRLISCFILNFMFYIGSNASRGRRQRRMHPNVSQCGGCATCSDYNGCLSKPR 60
Qy 61 LFFALERIGMKQIGVCLSSCPGYYGTRYPDINKTKKADCDTCFNNKFTCKKSGFYL 120
Db 61 LFFVLERIGMKQIGVCLSSCPGYYGTRYPDINKTKKADCDTCFNNKFTCKKSGFYL 120
Qy 121 HLKCLDNCPEGLRANNTMTCVSIHVCESEWNPSPCTKKGTCGFGCTETRVREII 180
Db 121 HLKCLDNCPEGLRANNTMTCVSIHVCESEWNPSPCTKKGTCGFGCTETRVREII 180
Qy 181 QHPSAKGNLCPPTNTRKCTVQRKKCKGKGRERKRRK 222
Db 181 QHPSAKGNPVPNQRDNLYSTKKEVFKGRARKKGRKTKK 222

RESULT 5
ID Q9Z132 PRELIMINARY; PRT; 265 AA.
AC Q9Z132;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Thrombospondin type 1 domain.
GN RSPONDIN OR R-SPONDIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RP Kamata T., Katsube K., Michikawa M., Yamada M., Mizusawa H.;
RT "R-spondin, a novel thrombospondin type 1 domain gene, expressed in
RL the dorsal neural tube.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016768; BAA75640.1; --
DR MGD; MGI:2183426; Rspndin.

```

```

DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR Pfam; PF00090; tsp_1; 1.
DR SMART; SM00261; FU; 2.
DR PROSITE; PS00092; TSP1; 1.
DR PROSITE; PS00092; TSP1; 1.
SQ SEQUENCE 265 AA; 29331 MW; FFE88964743F5963 CRC64;

Query Match 42.6%; Score 644; DB 11; Length 265;
Best Local Similarity 46.5%; Pred. No. 5.9e-51;
Matches 119; Conservative 36; Mismatches 77; Indels 24; Gaps 6;

Qy 5 LISWLFILNFMFYIGSNASRGRRQRRMHPNVSQCGGCATCSDYNGCLSKPRLPPA 64
Db 11 VLSWTHIA-----VGSRGI-KGRQRRIISAGSQACAKGELCSVNGCLSKPRFIL 63
Qy 65 LERIGMKQIGVCLSSCPGYYGTRYPDINKTKKAD-CDTCFNNKFTCKKSGFYHLG 123
Db 64 LERNDIRQGVCLPSCPYPGYFDARNPDNWKIKCKIEHCEACFSHNPCTKCBALYHLG 123
Qy 124 KCLDNCPEGLRANNTMTCVSIHVCESEWNPSPCTKKGTCGFGCTETRVREIIQHP 183
Db 124 RYPACPEGSTAANSITWEGSPACQCESEWSPGFCCKRKLQGRKGSERTRRVLHAP 183
Qy 184 SAKGNLCPPTNTRKCTVQRKKCKGKGRKGGK-----RERKKKPNKGSKBAPDSKSLB 239
Db 184 GGDHTTCSDTETKTRCTVRRTPCPGQRRKGGQRRRNNRHPARVNSKE--PRSNS-- 239
Qy 240 SSKEIPBQRENKQOKK 255
Db 240 -----RRHKGGQQ 247

RESULT 6
Q8N7L5 PRELIMINARY; PRT; 236 AA.
ID Q8N7L5
AC Q8N7L5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ40906.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RP TISSUE=Uterus;
RC Oshima A., Takabaishi-Pujii A., Tanase T., Imose N., Takeuchi K.,
RA Arita M., Muraashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Katsumi H., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahara K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK098225; BAC05263.1; --
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR Pfam; PF00090; tsp_1; 1.
DR SMART; SM00261; FU; 2.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS00092; TSP1; 1.
DR Hypothetical protein.
SQ SEQUENCE 236 AA; 25942 MW; 8D03803127EC5678 CRC64;

Query Match 41.5%; Score 626.5; DB 4; Length 236;
Best Local Similarity 48.1%; Pred. No. 2.1e-49;
Matches 111; Conservative 36; Mismatches 69; Indels 15; Gaps 3;

```


Db 1 MRCLPSFALLIILNCMDYSQCQ-GNRWRNRKRA-SYVSNPICKGCLSCSGKNGCSRQCK 58

Qy 61 LFFALERIGMKQIGVCLSSCPGSGYGYTRYIPDINKTKCKAD-CDTCFNKNFCTCKSGPY 119

Db 59 LFFFLREGMRQYGECLHSCPSGYGHRAPDMNRKARCIENCDSPSKDCTCKCKGVF 118

Qy 120 LHLGKCLDNCPSGLEANNHTWECVSIHVCEVSEWN 154

Db 119 LHRGRCDCEPCDGPAPILDETMCVB--GCBVGHWS 151

RESULT 11

Q8BJ73 PRELIMINARY; PRT; 138 AA.

ID Q8BJ73

AC Q8BJ73

DT 01-MAR-2003 (T-EMBLrel. 23, Created)

DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)

DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)

DN D3824P16.3.

GN A930029K19RIK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Retina;

RX MEDLINE=22354683; PubMed=12466851;

RA The PANTOM Consortium,

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs."

RL Nature 420:563-573 (2002).

DR EMBL; AK020904; BAC25643.1; -

DR MGD; MGI:1924467; A930029K19RIK.

DR InterPro; IPR006212; Furin.repeat.

DR InterPro; IPR009030; Grow_fac.recep.

DR SMART; SM00261; FU; 2.

SQ SEQUENCE 138 AA; 15172 MW; FEFD7D949279D5DD CRC64;

Query Match 18.8%; Score 283.5; DB 11; Length 138;

Best Local Similarity 42.9%; Pred. No. 2.9e-18;

Matches 45; Conservative 24; Mismatches 35; Indels 1; Gaps 1;

Qy 28 RRQRMRHPNVQSCQGGCATCSYNGCLSKTPRLFPALERIGMKQIGVCLSSCPGSGYGT 87

Db 22 RRKQAGTGLGGNCT-GCVTCSBENGSTCQQRLLFIRREGIRQYKGVHDCPLGPGFI 80

Qy 88 RYPDINKTKCKADDTCFNKNFCTCKSGFYHLHGLCLDNCBPG 132

Db 81 RGQBANRCKKOGATCSBSCFSQDFCIRCKRFRFLYKGRCLPSCP TG 125

RESULT 12

Q26489 PRELIMINARY; PRT; 1299 AA.

ID Q26489

AC Q26489;

DT 01-NOV-1996 (T-EMBLrel. 01, Created)

DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)

DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)

DN Endoprotease furin.

GN FURIN.

OS Spodoptera frugiperda (Fall armyworm).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;

OC Noctuidae; Amphipyrinae; Spodoptera.

OC NCBI_TaxID=7108;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Ovary;

RA Cleplik M., Klenk H.;

RT "cloning and functional characterization of FURIN from Spodoptera

RT frugiperda (sf9) cells."

```

RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC -I- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
DR EMBL; Z68888; CAA93116.1; -.
DR PIR; T43251; T43251.
DR HSSP; Q99405; IMPT.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR006212; Furin_repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR InterPro; IPR00209; Peptidase_S8.
DR InterPro; IPR002884; Peptidase_S8B.
DR InterPro; IPR009020; Protease_inhib.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF01483; P_protein; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR ProDom; PD000717; P_domain; 1.
DR SMART; SM00261; FU; 10.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolyase; Protease; Serine protease.
SQ SEQUENCE 1299 AA; 142020 MW; 4C3799C7B8C572AB CRC64;

Query Match 12.7% Score 192; DB 5; Length 1299;
Best Local Similarity 27.8%; Pred. No. 6.6e-09;
Matches 63; Conservative 26; Mismatches 78; Indels 60; Gaps 12;

QY 36 NVSQGCGGCGCATCD-YNGCLSKPRLFFALERIGMKQIGVCLSSCPGSGYGYTRYDPINK 94
DB 795 SVCRPCAACATCSERADGCTCEHLL-----VLHDTGCMASCPSPHYET---EDDM 843
QY 95 CTCKADCTCF--NKNFTCKKSGFYHLGKLDNCPGLEANNHTWECVSIHVCEVSE 152
DB 844 CAKCHESCDQCGPGETQCTCHPSTYALDGRVTSPPAYYADKKKEC---MRCFVG- 899
QY 153 WNPWSPCTKKGKTCGPRGTETRVREIIQHPSAKGNLCPP---TNETKRC-TVQRKKCQK 208
DB 900 ---CSTCT-----SAFLCEFWELNKKGKCMFVGSDKCSA 933
QY 209 GERGKGRERKRPKNK-----GESKE---AIPDSKSLSSKEIPE 246
DB 934 GEFAV---DQKCKRCNPACDCYGENEGHCLTCFNPLLQDYKCPVE 977

RESULT 13
Q8AY18
ID Q8AY18 PRELIMINARY; PRT; 913 AA.
AC Q8AY18;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Serine protease PCS-A.
OS Rana esculenta (Edible frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Rana;
OX NCBI_TaxID=8401;
RN [1]
RP SEQUENCE FROM N.A.
RA Gannon F., Jegou S., Vallarino M., Vieau D., Vaudry H.;
RT "Molecular characterization of the cDNA and localization of the mRNA
RT encoding the prohormone convertase PC5-A in the European green frog.";
RL J. Comp. Neurol. 0:0-0(2002).
DR EMBL; AY131292; AAN10146.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0004714; F:transmembrane receptor protein tyrosine kin. .; IEA.
DR GO; GO:0006118; F:electron transport; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.

DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. .; IEA.
DR InterPro; IPR003435; CytC_heme_BS.
DR InterPro; IPR006211; Furin-like.
DR InterPro; IPR006212; Furin_repeat.
DR InterPro; IPR009030; Grow_fac_recep.
DR InterPro; IPR00209; Peptidase_S8.
DR InterPro; IPR002884; Peptidase_S8B.
DR InterPro; IPR009020; Protease_inhib.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00882; Peptidase_S8; 1.
DR Pfam; PF01483; P_protein; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR ProDom; PD000717; P_domain; 1.
DR SMART; SM00181; EGF; 4.
DR SMART; SM00261; FU; 5.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Protease.
SQ SEQUENCE 913 AA; 101864 MW; 6C9020632C47D98B CRC64;

Query Match 11.9% Score 179.5; DB 13; Length 913;
Best Local Similarity 29.4%; Pred. No. 6.4e-08;
Matches 52; Conservative 22; Mismatches 66; Indels 37; Gaps 11;

QY 43 GGC-ATCSDYNGC-----LSCKRLFPALRIGMKQIGVCLSSCPGSGYGYTRYDPINKCT 96
DB 632 GPCDAECSDV-GCDGPGPHHCNCLHFFYK--AKNNTKICVSDCPGLGYPA---DKKCK 685
QY 97 KCKADCTCF--NKNFTCKKSGFYHL--LGKLDNCPGLEANNHTWECVSIHVCEVSE 152
DB 686 KCFTCTCLGSRSDQSCSKSYLNEETNSCVNCPDGYLNDKRLC----- 735
QY 153 WNPWSPCTKKGKTCGPRGTETRVREIIQHHP-SAKGNLCPPPTNET-RKCTVQRKKCQ 207
DB 736 ----KKCHENCKIC-----TSARICTEHLGLSLQGSTCAVTCEDGRYSVAVKKECE 783

RESULT 14
Q964D2
ID Q964D2 PRELIMINARY; PRT; 1101 AA.
AC Q964D2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Gal/GalNac lectin Ig11.
GN IGL1.
OS Entamoeba histolytica.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=5759;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HM1:IMSS;
RX MEDLINE=21391855; PubMed=11500468;
RA Cheng X.J., Hughes M.A., Huston C.D., Loftus B., Gilchrist C.A.,
RA Lockhart L.A., Ghosh S., Miller-Sims V., Mann B.J., Petri W.A. Jr.,
RA Tachibana H.;
RT "Intermediate Subunit of the Gal/GalNac Lectin of Entamoeba
RT histolytica Is a Member of a Gene Family Containing Multiple CXXC
RT Sequence Motifs.";
RL Infect. Immun. 69:5892-5898(2001).
DR EMBL; AF337950; AAK92361.1; -.
DR GO; GO:0005529; F:sugar binding; IEA.
DR GO; GO:0007157; P:heterophilic cell adhesion; IEA.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR009030; Grow_fac_recep.
DR PROSITE; PS01186; EGF_2; 1.
KW Lectin.
SQ SEQUENCE 1101 AA; 119512 MW; C8B6F5C8DB565A6C CRC64;

```

